Multivariate analysis of refiner operating data from a TMP newsprint mill

By R.P. Harrison, R. Leroux, P.R. Stuart

Abstract: Process values upstream and downstream of the primary and secondary refiners at a TMP-newsprint mill were obtained for a period of 34 months, and two types of multivariate analyses were performed: Principal Component Analysis (PCA) and Projection to Latent Surfaces (PLS). Pulp throughput dominated the results even within a relatively narrow range of normal production rates, and a first attempt was made at a physical interpretation of the model.

Any external factors such as seasonal variations and changes in incoming chip quality are beyond the control of the TMP mill operator. These could conceivably be compensated through controllable internal factors, but key pulp characteristics that could serve as quality targets are not always measured directly.

One option is to model in real time those parameters that cannot be measured continuously, in order to apply inferential control (“soft sensors”) as reported by Strand [1] and elsewhere [2, 3]. Before proposing any such control strategy, however, it is necessary to understand the correlations and trends that are inherent to the refining operation using historical data.

The TMP newsprint mill under investigation has had a high-speed PI data historian in place for 34 months, into which virtually all process and operating data for the mill are fed. The mill has over 6,000 data tags, some of which are updated every 10 seconds, representing millions of values per day.

The resulting data explosion has created a daunting mass of information, one for which the automated pattern-recognition techniques of multivariate analysis (MVA) are perfectly suited. Mill personnel have tried to establish relationships between the process variables by considering only a few at a time — a frustrating task at best, hence their interest in considering a new approach.

Previous papers applying MVA to paper production processes have shown that dozens or even hundreds of process parameters can be boiled down to a mere handful of underlying “latent variables,” corresponding to unmeasured fundamental characteristics of the system [1, 4-8]. Using data from the refiner section of a TMP newsprint mill, Browne [9] found that a portion of the variability in the pulp properties could be explained by two latent variables; the first related to wood freshness, as measured by the potential brightness of the wood, and the second to the ease with which energy could be applied to the wood, typically higher in summer.

The primary purpose of this paper is to perform MVA on daily average process data at this mill, to determine how many latent variables are required to adequately describe the refining section.
all picture. Periods of unusually low production (<100 t/d) were excluded beforehand, as previous experience had shown that these systematically produce major outliers.

Partial Least Squares (PLS) analysis, a form of multivariate regression, was performed using two medium fibre length fractions as the Y-set, and upstream data as the X-set.

CHARACTERISTICS OF INCOMING CHIPS

The mill has a maximum two-day chip pile inventory, so that there is minimum buffering of incoming chip variability. During the 34-month period under study, chips came from a variety of outside suppliers, and often the exact source and species were unknown.

Chip characteristics are measured on grab samples from every incoming truck shipment, and on instantaneous grab samples taken every eight hours at the TMP feed conveyor that feeds all four refining lines. Size distribution is measured with a Gradex unit at the chip-receiving lab; other parameters are measured manually.

PCA on the TMP feed conveyor data yielded a combined \( R^2 \) of 37% and \( Q^2 \) of 20%. The latter is the predictive power of the model, calculated by systematically removing a portion of the dataset and comparing to the model obtained from the remaining data.

For the chip data, only one principal component very strongly correlated with season:
- Higher in summer: chips longer than 5/8-in.
- Higher in winter: chips shorter than 3/8-in.; chip density; chip moisture.

These trends were already apparent from the raw data, and make sense physically; the interesting point is the lack of any other latent components, indicating no other statistically significant correlations between the chip properties. This may be partly due to the sparseness of the data, with only three grab samples per day.

Table I. Variables used in initial PCA.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>52FC16.5.P</td>
<td>Hydrosulphite consumption (L/m)</td>
</tr>
<tr>
<td>52XAI1.30.A</td>
<td>Specific energy — Line 1 (kWh/t)</td>
</tr>
<tr>
<td>Pex_L1.Blh</td>
<td>Brightness at No. 1 Latency Chest</td>
</tr>
<tr>
<td>Pex_L1.Cons</td>
<td>Consistency at No. 1 Latency Chest (Log)</td>
</tr>
<tr>
<td>Pex_L1.CSF</td>
<td>Freeness at No. 1 Latency Chest</td>
</tr>
<tr>
<td>Pex_L1.LMF</td>
<td>Avg. Fibre Length at No. 1 Latency Chest</td>
</tr>
<tr>
<td>Pex_L1.P2000</td>
<td>Fines Fraction at No. 1 Latency Chest</td>
</tr>
<tr>
<td>Pex_L1.R100</td>
<td>R100 Fraction at No. 1 Latency Chest</td>
</tr>
<tr>
<td>Pex_L1.R14</td>
<td>R14 Fraction at No. 1 Latency Chest</td>
</tr>
<tr>
<td>Pex_L1.R28</td>
<td>R28 Fraction at No. 1 Latency Chest</td>
</tr>
<tr>
<td>Pex_L1.R48</td>
<td>R48 Fraction at No. 1 Latency Chest</td>
</tr>
<tr>
<td>CopDENS</td>
<td>Chip Density at TMP Feed Conveyor</td>
</tr>
<tr>
<td>CopSICC</td>
<td>Chip % Solids at TMP Feed Conveyor</td>
</tr>
</tbody>
</table>

Note that chip brightness was not measured. Little or no information was available on the wood species, freshness of the wood, the age of the tree, chipper type or other factors which can impact on final pulp quality [8, 10]. This may account for the low predictive power of the model.

FINAL PAPER QUALITY

Final paper quality is of great importance to the mill, and can be influenced by many factors. In particular, Saltin [6] found that fibre length is a critical factor to tear strength.

This conclusion was supported by PCA analysis of hourly data from the mill, taken during a typical week in September 2002 when the chip supply was relatively constant. Comparison of pulp quality variables and newsprint quality variables showed that the proportion of medium-sized fibres strongly correlated with important paper characteristics such as permeability, stretch, burst strength and tear strength.

Based on these results, the R28 and R48 fibre length fractions were retained as the Y-set for PLS analysis at the refiner section. Note that no daily pulp handsheet test data were available for the individual refining lines during the 34-month period.

PCA ON SELECTED REFINER OPERATING DATA

An initial PCA was performed using only 14 of the 110 available variables, to gain insight into the over-all process, see Table I. The list was based on one used in a previous study [9].

The PCA yielded two principal components, with a combined \( R^2 \) of 44% and \( Q^2 \) of 24%. There were no major outliers, so all data points were retained in the model.

Figure 1 is a score plot of the observations, in this case daily averages. The ellipse indicates the 95% confidence limit. The first component strongly correlated with the time of year, in terms of average outdoor temperature and calendar seasons, with some segregation between the three different years under study. This component also strongly correlated with bleach consumption (higher in winter) and pulp brightness (higher in summer). It appears, therefore, that this component may be related to wood freshness, as reported by Browne [9]. The outliers in the top-right corner correspond to observations with higher fines and lower long-fibre fractions.

Figure 2 is the corresponding loadings plot, showing how the variables correlate with respect to the same first two components. The second component was dominated by fines fraction and fibre length, and to a lesser extent with Line 1 specific energy (these variables are in the dashed circles). As no handsheet testing data were available, it is not possible to relate this directly with strength properties, and so it is difficult to conclude whether this corresponds to the second component reported by Browne [9], despite the obvious similarities.
PCA ON ENTIRE REFINER DATA SET

The inclusion of all 110 tags led to a PCA with several major outliers, all of which were associated with periods of low production. When all data points with a production level below 150 t/d were removed, these outliers disappeared. Incidentally, the study of low production days may merit separate investigation in the future, as these correspond to start-up and shut-down days.

The final PCA yielded four principal components, with a combined $R^2$ of 44%, roughly the same as the previous case. The $Q^2$ value is much improved, however at 37%. This means that the model fit is about the same as before, but the predictive power is much better when using a large number of variables. This could be an advantage when creating a soft sensor, in order to give the controller the best chance of success.

The score plot shown in Fig. 3 shows the relationship between the first two components and time. Individual observations vary both between seasons and within a given season. Further investigation revealed that the individual seasons were segregated, for instance, the winters of 2000, 2001 and 2002.

The loadings plot in Fig. 4 shows a possible interpretation of these same first two components. Even though the low production points had previously been removed, pulp throughput continued to dominate the PCA results. This is probably due to the large number of variables that change when the throughput changes, such as dilution flows, screw feeder motor voltages and so forth. It may be necessary in the future to give these variables a lesser weighting in the model (at present all variables have been given equal weighting).

The second and third components are very similar to the two found earlier using the short list of 14 variables, although it is difficult to say with certainty if they are the same. The fourth component is less obvious, but appears to be related to refiner plate gap. All these possible interpretations would have to be verified either experimentally or through further investigation.

PLS ON MEDIUM FIBRE-LENGTH FRACTIONS

The complete list of variables was again used in performing PLS analysis, with $R^2$S and $Q^2$S for $Y$s. All other response variables were excluded. Three components were found, with a combined $R^2$ of 45% and $Q^2$ of 40%. The PCA results for the remaining ninety-odd $X$ variables strongly resembled those for the entire dataset.

The “variable importance plot” in Fig. 5 ranks the different $X$s in terms of importance. For instance, the first bar on the left, representing season, clearly dominates, followed by two other parameters: dilution water pressure and primary refiner plate position. The rest tend to decline in importance gradually, probably due to the very high degree of correlation between the $X$s.

In spite of the strong seasonal variability in the chip size, the different chip size fractions are not very prominent among the $X$s in the model. Other, more immediate operating factors in the refining section tend to dominate medium fibre length. However, the available chip characteristics were quite limited, with no information whatsoever about wood species, so it may simply be that chip quality is under-represented in the model.

It is important to emphasise that in the absence of a designed experiment, it is scientifically impossible to assign cause-and-effect relationships to the PLS results. It may simply be that certain $X$s trend with the $Y$s because both are being affected by a third, hidden factor.

CONCLUSIONS

Although statistically significant PCA models were obtained from the mill data, it seems that the available chip data do not tell the whole story. As a result, chip quality is probably under-represented with respect to refiner operating parameters.

Even though data from low production days were removed, pulp throughput dominated the PCA results. This occurs even within a relatively narrow range of production rates. The reason for this is the large number of process variables that are directly or indirectly related to throughput.

Other than throughput, the inherent components that were obtained for the system appear to relate to:

- Summer vs. winter, bleach consumption and pulp brightness;
- Fines fraction, fibre length, and specific energy; and
- Refiner plate gap.

Using all the available data appears to have some advantages in terms of the predictive power of the model. However, groups of variables that are highly correlated tend to dominate the model, such that it may be necessary to give them a lesser weighting in the future.

LITERATURE


Résumé: Les valeurs de fabrication en amont et en aval des raffineurs primaires et secondaires d’une usine de papier journal à base de PTM ont été mesurées durant 34 mois consécutifs. Deux types d’analyses multivariables ont été effectuées sur des moyennes quotidiennes, l’analyse en composantes principales (ACP) et la projection aux surfaces latentes (PSF), à l’aide de diverses combinaisons et variables. Bien que des modèles statistiquement significatifs aient été dégagés des données de l’usine, il semble que les données disponibles sur les copeaux ne racontent pas tout. La capacité de production de la pâte a dominé les résultats, même à l’intérieur d’une échelle relativement petite des taux de production normaux. Néanmoins, il a été possible de tenter une première interprétation raisonnable de la physique du modèle.


Keywords: NEWSPRINT, THERMOMECHANICAL PULP, REFINERS, PROCESS VARIABLES, DATA PROCESSING, STATISTICAL ANALYSIS, MODELS.