Multivariate analysis of seasonal pulp quality variations in a TMP mill

By T. Browne, K. Miles, D. McDonald and J. Wood

Abstract: Factor analysis of wood furnish and TMP pulp quality data identified two wood-related factors explaining pulp quality variations. These were wood freshness, indicated by potential brightness, and the ease of application of energy to the wood, indicated by higher tear strength. Wood freshness, giving higher brightness and lower bleach chemical consumption was higher in winter periods, probably because of wood storage conditions. Easier application of energy occurred in the summer.

This report describes the multivariate analysis of seasonal wood chip, pulp quality and specific energy data obtained from an Eastern Canadian TMP mill. Data available consisted of monthly averages for an extensive set of process and quality variables covering a period of three years. The goal was to identify seasonal wood or chip property variations which could be the source of seasonal variations in pulp properties, such as those observed by Tyrväinen [1] or Fuhr et al. [2]. To this end, factor analysis was used to relate chip quality, TMP operating data, pulp properties and season.

FACTOR ANALYSIS

Factor analysis is a multivariate analysis technique used to identify underlying patterns in highly correlated data sets [3]. The method separates variables in a data set into subsets, each consisting of well-correlated variables, but which are relatively independent of variables in other subsets. Each factor is then a linear function of the variables in a given subset; the coefficients of these functions may shed some light on underlying physical characteristics of the data set. The basic procedure is well described in standard statistics textbooks [4,5]; its application in the pulp and paper industry is discussed in Ref. 6, 7.

Factor analysis is well-suited to data sets with strong interrelations between the variables. It is most useful when combined with knowledge of the underlying physical behaviour of the process.

OVERVIEW OF WOOD YARD AND PULP MILL OPERATIONS

Data for this study was supplied by an Eastern Canadian TMP newsprint mill. During the period covered by the data, pulp was made from logs chipped onsite. The log supply was nominally a mix of 60% balsam fir and 40% black spruce, although spruce consumption levels varied seasonally from 35% to almost 50%. Throughout the months following the end of the spring thaw, fresh wood deliveries to the mill gate were generally directed to the chippers, with excess going to an onsite log pile. The pile served to ensure a constant supply to the mill when deliveries of fresh wood were interrupted at night or on weekends. Wood supplied to the chippers at this time of year was therefore largely fresh and the age of logs in inventory was generally low. Beginning in the fall, the inventory of cut wood was built up both in the mill’s wood yard and in roadside piles in the woodlands. In winter, this inventory of fall-cut wood made up a larger portion of the supply to the chippers, increasing to 100% of the feed during the spring thaw when logging roads were impassable and fresh wood was unobtainable. Wood supplied to the chippers at this time of year came largely or entirely from inventory, and the age of the logs from inventory was typically higher than during the summer and fall months. Wood age and species were thus both seasonal variables, with species being easier to quantify since the age of individual roadside piles was not recorded.

Chips were saved in storage bins with maximum residence times of a few hours; there were no chip piles. Chips were therefore fresh when refined, and chip age did not vary seasonally. Following washing and steaming, chips were refined in the TMP plant. After screening, rejects were mixed with rejects from paper machine screening and cleaning operations, then processed in the rejects refining system. After further screening and cleaning, accepts from the rejects system were mixed with accepts from the mainline refiners at the disc filters. Following bleaching with hydro-sulphite, the pulp was mixed with excess paper machine white water in a blend chest. Pulp properties reported here were obtained from samples taken at this chest, and thus represent a composite of the entire TMP plant diluted with a certain amount of white water and fibre which has been rejected by the paper machine screens, then re-refined in the rejects system.

SELECTION OF VARIABLES

The analysis was performed using monthly average data. These monthly averages were available for a period of 35 months, from January 1995 to November 1997. A smaller selection of weekly data was also available and is described in Ref. 7.

Visual inspection of the monthly data reveals a steep drop in tear 7, Fig. 1, in the summer of 1996. This was largely due to changes in testing and reporting methods and does not reflect true
changes in pulp properties. There was also a drop in the long fibre fraction, or L-factor Fig. 2, caused by a drop in each of the R14, P14-R28 and P28-R48 fractions. The R14 fraction tended to drop every summer while the other fractions showed a less pronounced seasonal trend Fig 3. The R14 fraction was thus retained as a more reliable indicator of fibre length than L-factor. Tear was retained in spite of the step change because of its importance as a pulp strength indicator and because the accompanying decrease in fibre length, a known cause of reduced tear, suggested that not all of the step change in tear was due to procedural changes in testing.

As seen in Fig. 4, chip moisture content showed a clear seasonal trend, reaching a peak in late winter. Spruce content Fig 5. also reached its maximum value in the late winter or early spring. Specific energy and freeness showed less obvious seasonal trends, while burst and breaking length were typically low in the summer months. Bleach consumption was lowest in late winter, corresponding to peak values of brightness Fig 6. The full list of variables initially selected for analysis is provided in Table I.

STATISTICAL ANALYSIS

In order to compare variables on an equivalent basis, all data was first normalized to a mean of 0 and a variance of 1. Two factors were then extracted from the set of normalized monthly averages. The methods used, which were principal components extraction followed by Varimax rotation, are described in the literature [4,5]. Burst and R14 were uncorrelated with either of the two factors. These variables were deleted from the data set and the analysis was repeated. The matrix of factor loadings is shown in Table II, where columns give the correlation coefficients between a given factor and each variable, while individual rows provide regression coefficients for predicting the value of a variable from values of the factors. Regression coefficients below 0.50 have been neglected, as they account for less than 25% of the variance in the particular variable. Coefficients over 0.70 (explaining 50% or more of the variance) have been highlighted. The last column gives the portion of the total variance of each variable which has been explained by all coefficients of the two factors, including those below 25%.

Factor 1 is highly correlated with spruce content, bleach rate and brightness: Factor 1 increases when spruce contents and brightness are high, and when bleaching consumption is low. Factor 1 is also correlated to a lesser degree with tear, bulk and freeness. High tear, low bulk and low freeness will also cause Factor 1 to increase. Factor 1 explains 32.5% of the variance in the data set.

Factor 2 is strongly correlated with chip properties, refining energy and freeness, and will be high when chips are denser and drier, and when refining energy or freeness are high. Factor 2 is also correlated to a lesser degree with high tear and high values of SEASON, i.e. with warm summer months. Factor 2 explains a further 31.6% of the variance in the data set.

Several measures can be used to estimate the quality of the fit of these factors to the data set. The cumulative total variance explained by two factors is 64.2%, not an unreasonable number when analyzing noisy mill data. The last column in Table II shows that two factors explain 33.3% of the variance in the variable SEASON, 84.3% of that in SPRUCE, and so on. The poorest level of variance explained, apart from the arbitrary variable SEASON, is for bulk at 46.8%. Tear is the variable which is least well explained by two factors while the overall variance explained is 65%, this is distributed across the two factors, with no one factor explaining more than 38% of its variance.

The variable SEASON may be poorly explained due to changes in spruce content and specific energy over the three-year period. There was a general trend over the three-year period to lower specific energy, while Fig 5 and 7 show a general decrease in spruce content from year to year. These overall shifts in specific energy and species mix may be reducing the portion of the variance explained by seasonality.

Browne et al [7] presented an error
DISCUSSION

Two fundamental principles must be kept firmly in mind when interpreting multivariate analyses of industrial process data. First, statistical analysis reveals only associations or correlations between variables; it cannot prove cause and effect. A large correlation between two variables does not in any way imply that a change in one must be the cause of the change in the other, or vice versa. The changes in both may be caused by a third variable not included in the data set. Secondly, in any industrial data set extending over time, actions by plant operators or automated control systems in response to changes in one variable can have an influence on the ease of application of energy in a refiner or chipping, genetics, etc. These independent variables have an effect. The result is higher energy use leading to well-fibrillated, better retention or brightness potential of the wood.

Weekly averages over a longer time span were also available, but for a smaller set of variables. Analysis of this data gave similar results to the monthly data [7]. These various measures of statistical fit imply that the model predicts the data reasonably well given the scatter inherent in mill measurements.

Pulp & Paper Canada T 232

analysis of the model as given in Table II, which shows that the model predicts all but one of the significant correlations at the 95% level.

Weekly averages over a longer time span were also available, but for a smaller set of variables. Analysis of this data gave similar results to the monthly data [7]. These various measures of statistical fit imply that the model predicts the data reasonably well given the scatter inherent in mill measurements.

First, statistical analysis reveals only associations or correlations between variables; it cannot prove cause and effect. A large correlation between two variables does not in any way imply that a change in one must be the cause of the change in the other, or vice versa. The changes in both may be caused by a third variable not included in the data set. Secondly, in any industrial data set extending over time, actions by plant operators or automated control systems in response to changes in one variable can have a profound effect on correlations of that variable with others, or between two others. Nevertheless, multivariate analysis provides a powerful guide for assessing possible causes of observed variations, or when designing experiments to prove cause and effect.

It is clear that high brightness pulps occur during periods of high spruce consumption as shown in Fig. 6, and in the factor loadings for Factor 1. It could be suggested that spruce pulps are inherently brighter, thus requiring lower levels of bleaching chemicals to maintain a minimum brightness level, but this is inconsistent with reports showing fresh fir can be as bright as fresh spruce [8]. We prefer to suggest that brightness loss during storage for either spruce or fir is reduced during the period when coincidentally, spruce content is high, in this case the cold weather period. This would explain the apparent counter-intuitive result that lower rates of bleach addition yield higher brightness pulps, as seen in Fig. 6 and 8. Factor 1 thus models the effect of wood freshness on brightness, and implies that improved wood freshness (due to shorter storage times, lower storage temperatures, etc.) will improve brightness and reduce bleach chemical consumption. It is a measure of brightness retention or brightness potential of the wood.

When the wood allows more energy to be put into the pulp, spruce consumption is high. However, we believe this is coincidental, that is, the cold weather period. This would explain the apparently counter-intuitive result that lower rates of bleach addition yield higher brightness pulps, as seen in Fig. 6 and 8. Factor 1 thus models the effect of wood freshness on brightness, and implies that improved wood freshness (due to shorter storage times, lower storage temperatures, etc.) will improve brightness and reduce bleach chemical consumption. It is a measure of brightness retention or brightness potential of the wood.

When the wood allows more energy to be put into the pulp, spruce consumption is high. However, we believe this is coincidental, that is, the cold weather period. This would explain the apparently counter-intuitive result that lower rates of bleach addition yield higher brightness pulps, as seen in Fig. 6 and 8. Factor 1 thus models the effect of wood freshness on brightness, and implies that improved wood freshness (due to shorter storage times, lower storage temperatures, etc.) will improve brightness and reduce bleach chemical consumption. It is a measure of brightness retention or brightness potential of the wood.

Factor 2 is also correlated with denser, drier chips, which tend to be more common in the summer when the variable SEASON is high. However, we believe this is coincidental, that is, the cold weather period. This would explain the apparently counter-intuitive result that lower rates of bleach addition yield higher brightness pulps, as seen in Fig. 6 and 8. Factor 1 thus models the effect of wood freshness on brightness, and implies that improved wood freshness (due to shorter storage times, lower storage temperatures, etc.) will improve brightness and reduce bleach chemical consumption. It is a measure of brightness retention or brightness potential of the wood.

At first glance the inclusion of freshness in both Factor 1 and
Factor 2 appears contradictory. It is not. Freeness is known to be influenced by three largely independent pulp properties, specific surface, how well the fibres can be compacted, and fibre length distribution in the short and medium fractions [11]. Apparently these individual contributors to freeness separate into different factors.

Statistical procedures like factor analysis do not prove cause and effect. They merely identify patterns in the data. A knowledge of the underlying physical meaning of the variables is essential for linking potential cause-and-effect relationships with the statistical correlations. With this in mind, factor analysis has been used with some success as the basis for pulp and paper quality control systems in a number of mills [12-15]. In this method, factors are used to identify the corrective actions necessary to make up for a loss in properties. The loadings in Table II can be used to calculate the value of the factors, as follows:

\[ F_i = \sum a_{ij} x_j \]

where Factor \( i \) is the sum of the \( j \) factor loadings for Factor \( i \), \( a_{ij} \) multiplied by the appropriate value of the normalized variable, \( x_j \). Using this study as an example, Factor 1 is the sum (– 0.2692 (SEASON) + (0.9183 \times SPRUCE) + (...). Ideally the mill should be operated so as to maintain the value of this factor. For example, if tear should drop, the value of the factor will also drop, and other variables, such as the level of spruce, can be adjusted to compensate. Assume for instance that tear drops by some fraction \( \alpha \), then one can calculate, in principle, what increase in spruce consumption \( \beta \) is necessary to maintain \( F_1 \) constant by means of the following expression:

\[ \beta = \alpha \frac{a_{1T}}{a_{1S}} \]

where \( a_{1T} = 0.5150 \) is the tear loading for Factor 1, and \( a_{1S} = 0.9183 \) is the spruce loading for Factor 1. The ratio is therefore 0.5608, which suggests that a 5% drop in tear could be offset by increasing the spruce content by 2.8%. Similar calculations can be made for other pairs of variables, with the understanding that the predictions are only valid over the range of values covered by the data. For example, maintaining the value of Factor 2 in the face of a 5% drop in tear could be achieved by increasing specific energy by 3.8%.

For this approach to work, there must be a cause and effect relationship between the measured variables and the factors. We have suggested possible cause and effect relationships which account for the correlations without contradicting the known effects of wood properties on pulp properties, but we caution that these relationships should be verified experimentally, for instance by deliberately changing spruce content without altering freshness or vice versa.

It also needs to be pointed out that the choice of variables to be included in a factor analysis model is critical. The number of factors to retain can also be a difficult choice: more factors explain more of the variance in the data set, but each new factor is more difficult to interpret than its predecessor. Factor analysis and other similar methods such as principal components analysis, have procedures for making these choices. In our experience however, these choices are essentially subjective based on prior knowledge of the process including the reliability of each variable. In interpreting the data available in this study, we repeated the analysis with several different combinations of variables; each combination was analysed using two, three and four factors. On the basis that the simplest explanation consistent with the facts is the most likely to be useful, we chose to present a two-factor model. Examples of three-factor models are given in Ref. 7.

**CONCLUSIONS**

In summary, factor analysis implies that pulp properties at this mill depend on wood properties in two different ways. A portion of the variation in pulp properties, identified as Factor 1, is related to the potential brightness of the wood. We consider this to be a measure of wood freshness: fresher wood, which has been stored for a shorter time, at lower temperatures, or in conditions less conducive to biological attack, tends to make a brighter pulp, requiring less bleaching chemicals to reach a given brightness target. In this mill during the period when the potential brightness and wood freshness are high, spruce content, sheet density and tear strength are also slightly higher than in other periods and freeness is somewhat lower.

Another portion of the pulp variability, identified as Factor 2, is due to physical properties of the wood, which we believe influence the ease with which energy can be applied to it. Tear strength is higher in the summer since higher energy levels can be applied to the wood used in the summer, without reducing freshness by excessive fibre cutting. Summer chips also tend to be drier and denser. Differences in the genetics of the wood adapted for growing in different areas, as well as the direct influence of each environment on wood growth, may in part be responsible for differences in the wood properties which influence how easily energy can be applied to it in a refiner. For example, winter wood is more likely to be cut in low-lying, marshy areas, and summer wood on hillsides and in mountainous areas. Other examples are possible. It is reasonable to conclude that genetic and environmental factors would produce a geographic variation in wood properties which could in turn be the cause of the observed differences in pulp properties.

A final conclusion, in accord with others [15], is that with the introduction over the last decade of quality and inventory control systems based on computerized data acquisition, and of systems for storing and indexing this data, more and more mills now have access to a large amount of historical data describing their
process. Especially where a mill-wide data collection system is also available, PC-based statistical software may be used to identify relations between process and product variables, which may not otherwise be apparent. In turn these relationships may be used to guide the planning of experiments to identify the causes of product quality variations.

The user should remain aware that these procedures inevitably require subjective interpretation based on a sound understanding of the process and that a verification step is essential to establish cause and effect.

REFERENCES