

Monitoring kraft recovery boiler fouling using principal component analysis

PETER VERSTEEG AND HONGHI TRAN

ABSTRACT: Researchers analyzed high resolution operational data from three recovery boilers using the principal component analysis (PCA) feature of a multivariate statistical analysis program to identify major operating variables that contributed to fouling and plugging. The results show that PCA can be used to visualize the variability related to long-term fouling trends in the boilers and to graphically distinguish changes in the boiler fouling condition caused by operational variability over a short period. This represents a major step forward in identifying operating variables that might be adjusted to minimize fouling, and in developing an on-line fouling monitoring technology based on PCA.

Application: Mills can use PCA to identify major operating variables that correlate with recovery boiler fouling and to devise appropriate control strategies.

Fouling, or the accumulation of fireside deposits on heat transfer tube surfaces, is a persistent problem in the recovery boilers of many kraft mills. Fouling drastically decreases the steam generation efficiency of the boiler, and in severe cases can plug the flue gas passages, which leads to unscheduled boiler shutdowns for deposit washing. The extent of fouling in a recovery boiler depends on many factors, including boiler design and operating capacity, black liquor composition and combustion behavior, boiler operating conditions, and soot-blowing efficiency [1]. These factors often are related to one another; it is therefore difficult to identify one or two factors that are primarily responsible for fouling and to devise a viable solution to the problem.

Principal component analysis (PCA) and partial least squares (PLS) analysis are two complimentary multivariate statistical techniques that have been used to identify the main causes of problems in various industrial processes [2,3], including in the pulp and paper industry [4]. These techniques also have been used to identify the main causes of ring formation in a lime kiln [5], and to help identify the main operating variables that cause fouling in two recovery boilers [6]. PCA allowed graphical and operational comparisons between a high fouling period and a low fouling period in both recovery boilers. We used PLS to extract the correlation structures between variables, provide a better understanding of overall boiler operation, and focus on which variables might be adjusted to improve boiler performance.

As an extension of previous work [6], the objectives of this study were to compare the “movement” or deviation of the data on PCA plots during periods when the boilers experienced severe fouling problems (high-fouling periods) to periods when the boilers were running well (low-fouling periods), and to determine if such movement can be used to monitor the onset and the extent of severe deposit buildup

(fouling) in the boilers on a day-to-day basis. This paper first describes the fouling experience of the boilers involved and the basic mathematical principles of PCA, and then discusses how PCA might be used to identify main factors that can contribute to the fouling problems and to monitor the boiler fouling condition with time.

FOULING EXPERIENCE

The fouling conditions of three recovery boilers, RB-A, RB-B, and RB-C, were investigated in this study. RB-A is a two-drum unit designed to burn 1.6 million kg/day of as-fired black liquor dry solids (BLDS) and to produce 250,000 kg/h superheated steam rated at 440°C and 6.2 MPa. The boiler experienced, on average, two plugging outages per year between 2002 and 2005. The experience was particularly bad in 2005, when water washing was a two-to-three month ritual. **Figure 1** shows a heavy deposit buildup at the generating bank inlet



1. Heavy deposit buildup at the generating bank inlet of recovery boiler A (RB-A).

of RB-A immediately before a water wash. The boiler had been operating for about two months. RB-B is at the same mill as RB-A. It is a single drum unit with a designed firing capacity of 2.4 million kg/d BLDS and a steam production capacity of 320,000 kg/h steam. The boiler experienced three water washes a year between 2002 and 2005. RB-C, located at a different mill, is a two-drum boiler designed to burn 2.0 millions kg/day BLDS and produce 275,000 kg/h steam at 470°C at 6.7 MPa. The boiler also experienced three water washes a year between 2004 and 2007.

DATA COLLECTION

Useful information might be contained in the correlations between variables; therefore, as many process variables should be collected as possible [7]. Common process variables for the three recovery boilers were identified using historical databases and laboratory records from the mills involved. For each boiler, variables included black liquor firing load, solids content, temperature, pressure, chlorine and sodium content, air flow rates (primary, secondary, and tertiary), air temperature and pressure, feed water flow rates and temperature, steam flow rate, temperature and pressure, soot-blowing steam consumption, flue gas pressure drops across the boiler bank and economizer, flue gas temperatures at the boiler bank outlet and economizer outlet, excess oxygen, induced draft fan speed, total reduced sulfur and sulfur dioxide emission data, and white liquor sulfidity. The data resolution for each variable ranged from hourly-averaged measurements taken by on-line instruments, to mill laboratory data collected as infrequently as a few times per year. Steam enthalpy was calculated based on steam flow rate, pressure, and temperature data.

PRINCIPAL COMPONENT ANALYSIS

In this study, PCA was performed using SIMCA-P (Umetrics AB, Umeå, Sweden), a commercially available multivariate data analysis program. PCA is a modeling technique that can be used to visually untangle the interactions among highly correlated variables in complex systems. Large amounts of process data from multiple variables can be represented as a single data point that moves in time along multiple axes called principal components (PCs). Principal components are new variables built using PCA; they represent statistically significant variable interactions in the system and are usually smaller in number than the original process variable set [7,8]. Each principal component models a portion of the variability in the data set. Because many instruments are used to measure a small number of physical system properties that cause variability, a handful of principal components often can be used to accurately represent the system.

For a data set extracted from an industrial process, the data are normalized to eliminate the effects of differing units, and averaged by mean centering the variables and scaling them to unit variance. The resulting data can be expressed in terms of the matrix X :

$$\underline{X} = [\underline{x}_1 \ \underline{x}_2 \ \underline{x}_3 \ \dots \ \underline{x}_r]$$

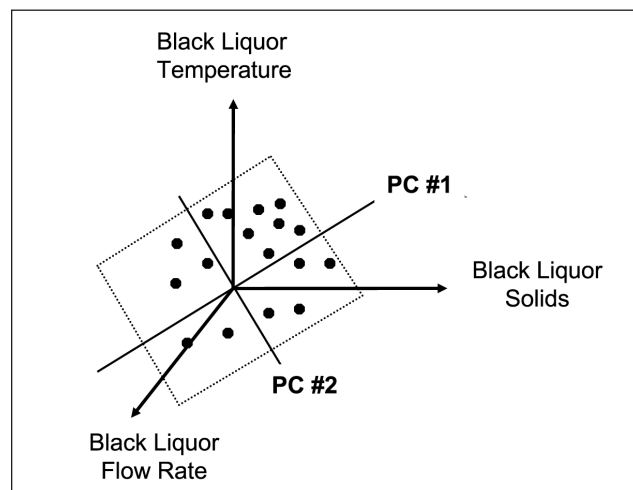
where $\underline{x}_1, \underline{x}_2, \underline{x}_3, \dots, \underline{x}_r$ are vectors of data measured by r on-line process instruments. PCA decomposes the normalized data matrix $\underline{X}$ into three new matrices ($\underline{T}$, $\underline{P}$, and $\underline{E}$), using the relation:

$$\underline{X} = \underline{T} \underline{P}^T + \underline{E}$$

where $\underline{T} = [t_1 \ t_2 \ t_3 \ \dots \ t_k]$ is defined as a principal component score matrix that contains the decomposed process data, $\underline{P} = [p_1 \ p_2 \ p_3 \ \dots \ p_k]$ is a principal component loading matrix that identifies variables that are associated with processes in the system, and $\underline{E} = [e_1 \ e_2 \ e_3 \ \dots \ e_r]$ is a residual matrix. Subscripts k and r refer to the number of principal components calculated and the number of on-line process instruments, respectively.

In many systems, the mathematical decomposition forming $\underline{T}$ and $\underline{P}$ can accurately approximate the original data set using only a few principal components. In these cases, most of the important process information is contained in smaller matrices, greatly simplifying the system analysis. These two matrices can be used to compare data clusters, identify process upsets or faults, and visualize the variability caused by specific processes such as deposit accumulation.

Figure 2 shows the mathematical decomposition visually. Each original process variable gives an independent axis in multidimensional space. Here, the black liquor solids, black liquor flow rate and black liquor temperature are shown on the x , y , and z axes, respectively. Any process measurement across all the variables can be represented as a single point along these axes. Subsequent simultaneous measurements of all the variables become additional data points and a cluster of data points appear in multidimensional space. Although many variables often are measured simultaneously through on-line instruments, the measured data for a particular instant can be represented as a single data point in this space.



2. Visual description of principal component analysis (PCA).

RECOVERY BOILERS

Because each data point can represent an entire day of data across all the variables measured, Fig. 2 can be used to “visualize” large amounts of data across extended periods. An additional line (PC#1) representing a principal component is drawn as a line through the data points. This first principal component is a linear combination of process variables weighted such that the line is in the direction of the largest variability in the data set. The combinations of variables with the highest weights that form this principal component represent the main physical process that drives the largest amount of variability in the data. A second principal component (PC#2) can be drawn so that it is orthogonal, or perpendicular in every direction, to the first principal component. This second principal component is in the direction of the next largest source of variation within the data set. Together, these two principal components form a window through which the data and the variables can be investigated. This window is called a t_1/t_2 plot, which can be used to visualize variability resulting from deposit buildup and to compare periods of high fouling with periods of low fouling.

PCA RESULTS

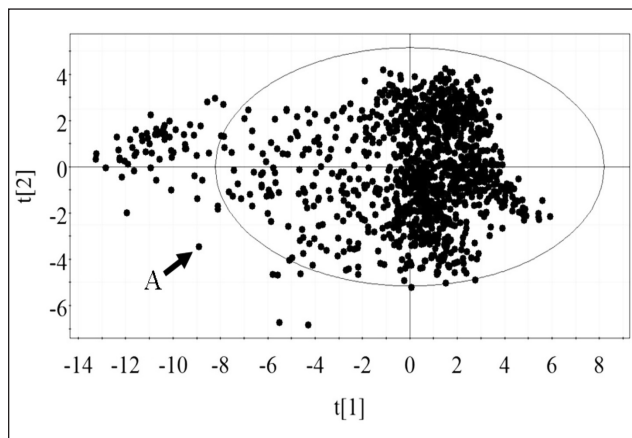
PCA models were built for all three recovery boilers using the method described in the SIMCA-P user guide [9]. PCA was performed on daily average data compiled during a period spanning three years and 10 month for RB-A and RB-B. For RB-C, PCA was performed on hourly average data for two years and six months so that particular runs could be investigated at a higher data resolution. In all cases, the data collected during the periods when the boilers were down for either water washes or other reasons was removed from the data sets. In general, black liquor chlorine and sodium data were not available for the specific days investigated in this study, and so these variables were excluded from the results.

The principle components for each data set were calculated through cross-validation, which first kept part of the data out of model development, then used the model to predict the data that have been left out, and compared them with the actual values [9]. The principle components accounted for approximately 69%, 66%, and 64% of the variability in the data for RB-A, RB-B, and RB-C, respectively.

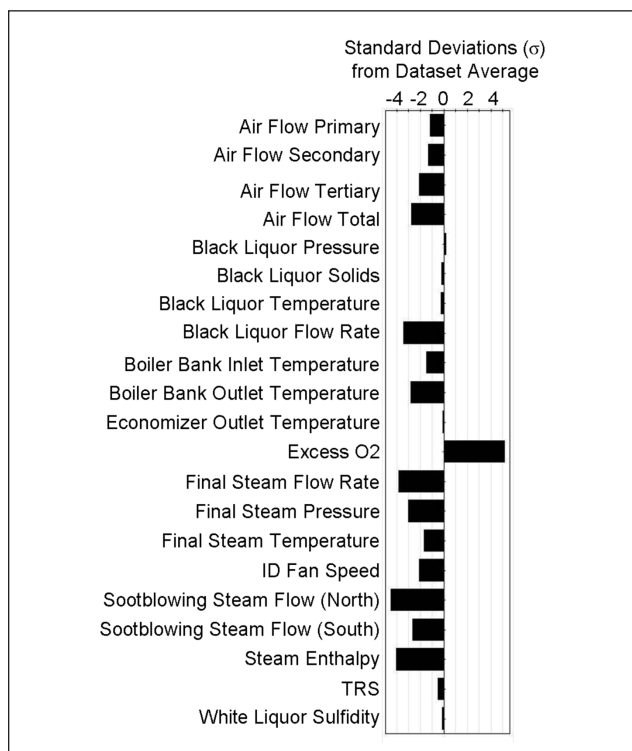
RB-A

Figure 3 shows the first and the second principal components plotted as the axes on the two-dimensional t_1/t_2 plot for RB-A. Each data point represents an entire day's worth of data across all the variables. The center (origin) of the plot corresponds to the combined average of all the data. Data points located away from the center indicate that these points deviate from the average. The ellipse is an area of 95% confidence interval, into which 19 of 20 data points fall.

Each data point in Fig. 3 can be analyzed individually in terms of the values of the variables that make up that data point in a contribution plot. For example, the contributions of variables to the position of point “A” are shown in **Fig. 4**,



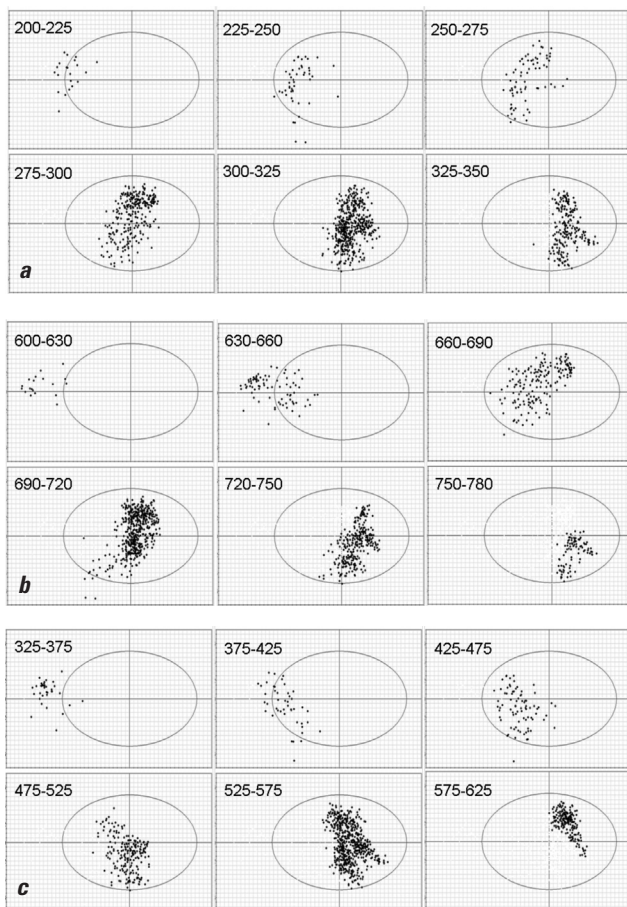
3. The t_1/t_2 plot for the RB-A model. Point “A,” used as an example, represents an entire day's worth of data across all the variables that were measured.



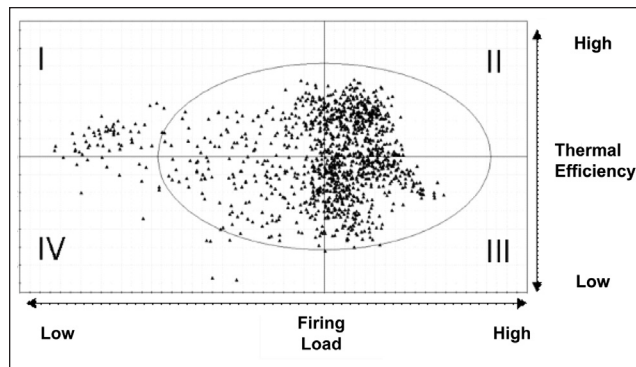
4. Contribution plot of variables for point “A” in Fig. 3. The horizontal axis represents standard deviations of each variable from the data set average, with 0 representing the average.

in terms of their values and differences from the data set averages. The 0 line in Fig. 4 represents the average of each variable across all the data collected for RB-A; most variables are at least one standard deviation from the average. The contribution plot suggests that point “A” was the day when RB-A was operated at a lower total air flow rate, a lower black liquor flow rate, a lower steam flow rate, a lower soot-blowing steam flow rate, a lower flue gas temperature, and at a higher excess oxygen than the data set average. These are characteristic of boiler operating conditions during a low liquor firing load.

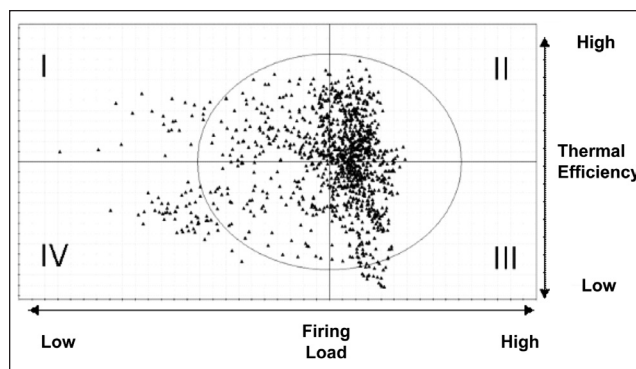
The data points in Fig. 3 were divided into separate plots of increasing values of black liquor flow rate, boiler bank outlet temperature, and final steam flow rate (**Fig. 5**). The results



5. Variation in the t_1/t_2 plot with respect to (a) black liquor flow rate, (b) boiler bank outlet temperature, and (c) final steam flow rate for recovery boiler A (RB-A).



6. Key processes in the t_1/t_2 plot of the PCA model for RB-A.



7. Key processes in the t_1/t_2 plot of the PCA model for recovery boiler B (RB-B).

show (Fig. 5a) that as the black liquor flow rate increases, the data points shift from left to right along the t_1 axis, indicating the variability of the first principal component with increasing black liquor firing load. As the boiler bank outlet temperature increases (Fig. 5b), the data points also shift from left to right along the t_1 axis, but concentrate mostly in the lower part of the t_2 axis. Similarly, as the final steam flow rate increases (Fig. 5c), the data points shift from left to right along the t_1 axis, but concentrate mostly in the upper part of the t_2 axis.

Because both increasing the boiler bank inlet temperature and decreasing the steam production are indicative of increased superheater fouling, the above results suggest that during the time when the thermal efficiency of the boiler is high (i.e., high firing, high steam production and lower boiler bank outlet gas temperature), the data points tend to be in the second quadrant (upper-right) of the t_1/t_2 plot; and as the fouling proceeds, they move toward the fourth quadrant (lower-right) of the plot. The concept can be explained using **Fig. 6**. The upper-right quadrant (II) contains data points that correspond to desirable operating days when the boiler was firing at a high load and had high thermal efficiency. The lower-right quadrant (III) shows data points for days when the boiler was at a high load, but had low thermal efficiency, presumably because of severe fouling. The upper- and lower-left quadrants (I and IV) represent operating days when the boil-

er was at low firing loads with high and low thermal efficiency, respectively.

Higher principal components, which accounted for less of the remaining variability in the data, also were calculated. The third through sixth principal components yielded weak correlations primarily involving boiler pressure, economizer temperature, tertiary air flow, black liquor solids, and black liquor pressure.

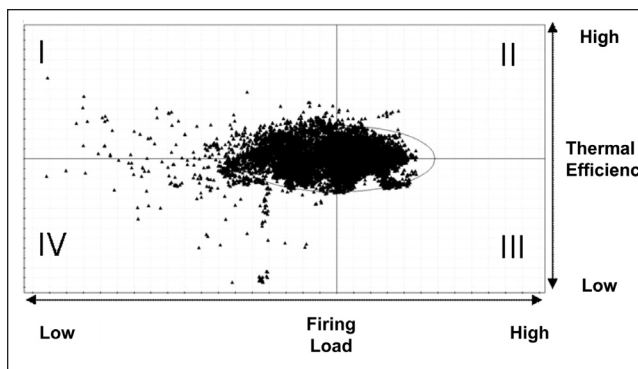
RB-B

Figure 7 summarizes the same two key processes for RB-B. Each data point is a representation of an entire day's worth of data across all the variables that were measured for RB-B. As with RB-A, the first principal component in this case explains the variability caused by boiler loading, while the second principal component explains the variability caused by changes in thermal efficiency. For both RB-A and RB-B, the data points concentrate in the II and the III quadrants of the t_1/t_2 plots, indicating that the boilers were operated most of the time at a high firing load, but at varying thermal efficiencies.

RB-C

For RB-C, sporadic instrument problems for some process variables reduced the availability of key temperature and flow measurements. As a result, hourly average data points were used, and thus, each data point in **Fig. 8** is a representation of an hour's worth of data across all the variables. There was

RECOVERY BOILERS



8. Key processes of the t_2/t_4 plot of the PCA model for recovery boiler C (RB-C).

also a major process change apparent in the data set. A large portion of the secondary air flow and a smaller portion of the primary air flow had been diverted to the tertiary air ports causing a large amount of variability in the data. As a consequence, the statistically significant processes representing boiler loading and thermal efficiency were found at the second and fourth principal components, respectively.

LOW FOULING VS. HIGH FOULING

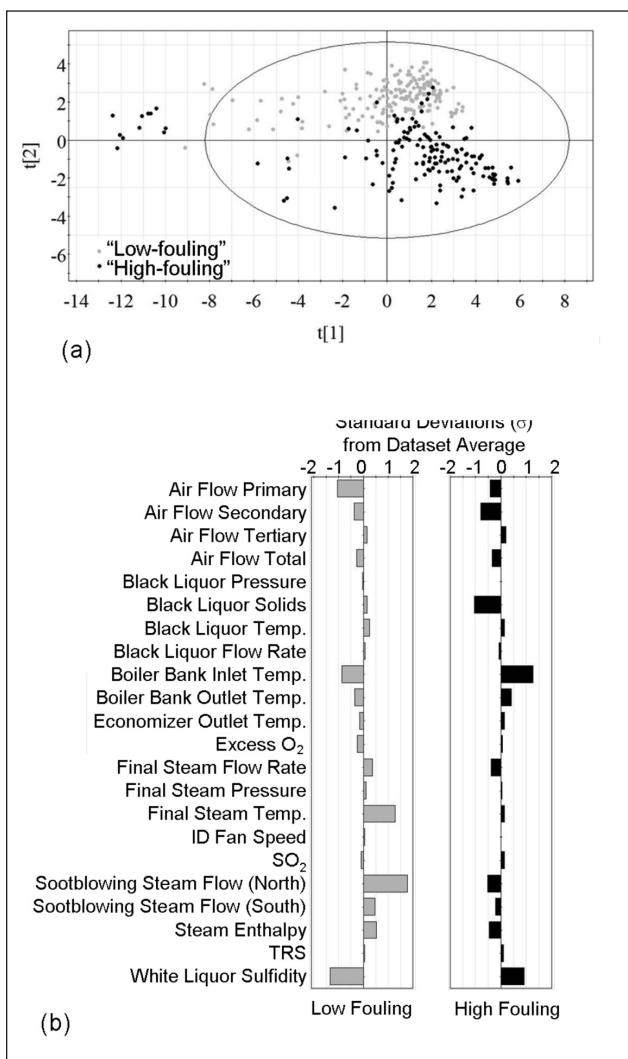
RB-A

The data set for RB-A was separated into several sections and the number of fouling-related water washes was determined in each section. Periods were compared based on the number of fouling-related water washes they contained. A seven-month “low-fouling” period began with a clean boiler and ended with a scheduled water wash. A five-month “high-fouling” period combined a two-month and a three-month period to encompass three water washes. The two periods were plotted over Fig. 6 with all the other data points removed. **Figure 9** shows the results and the corresponding contribution plot comparing the two periods.

In Fig. 9a, grey circles represent days corresponding to the “low-fouling” period and black circles represent days corresponding to the “high-fouling” period. The data points clearly cluster into two distinct regions. Data points for the “high-fouling” period are primarily situated at the bottom of the t_1/t_2 plot (the lower thermal efficiency region); the data points for the “low-fouling” period are situated almost entirely in the upper part of the plot (the high thermal efficiency region). Little overlap occurs between the two periods on the t_2 axis, suggesting that these periods were significantly different in terms of thermal efficiency. Figure 9b is a contribution plot comparing the averages of variables between the low fouling and high fouling periods. The 0 line represents the average of each variable across all the data collected for RB-A. Figure 9b shows that the operation with high white liquor sulfidity and low solids firing might have contributed greatly to this boiler’s fouling problem.

RB-B

Because RB-B experienced consistent fouling problems, it was difficult to divide and compare sections of the data set in terms

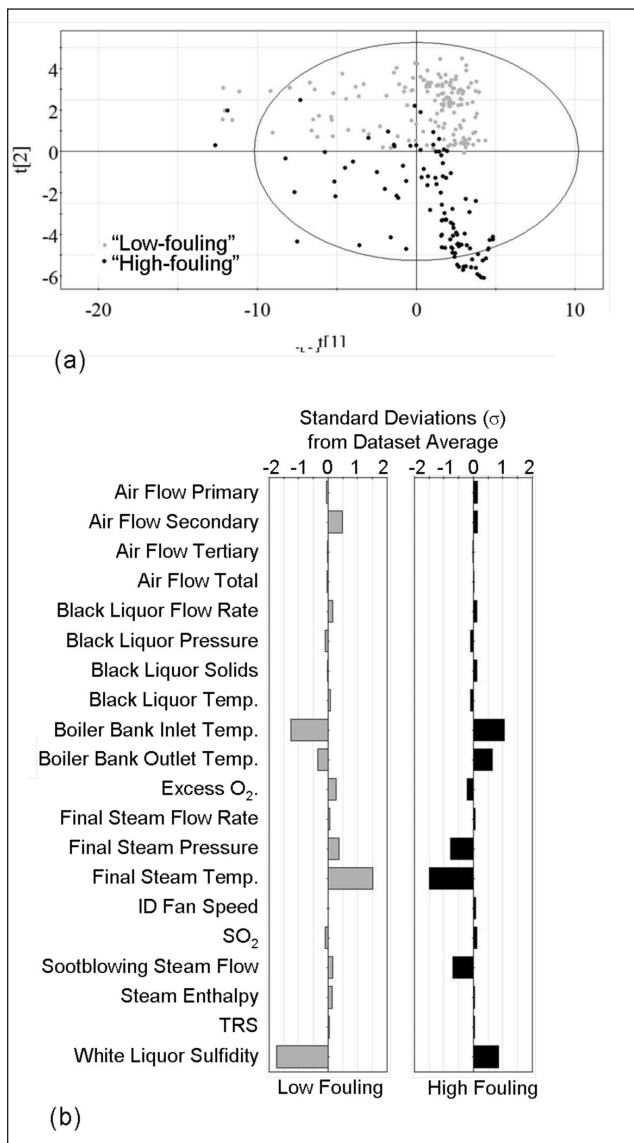


9. Comparison of low-fouling and high-fouling periods for RB-A and corresponding contribution plot.

of the number of water washes. Instead, “low-fouling” and “high-fouling” periods were selected by comparing fouling indicator variables such as the boiler bank outlet temperature. During some periods, the boiler bank outlet temperature rose sharply, indicating rapid occurrence of fouling. During other periods, the increase in boiler bank outlet temperature with time was not significant. The “low-fouling” period was from October 3 to March 14. The “high-fouling” period combines operational runs from March 3 to May 5 and from July 13 to September 5. The two periods were plotted over Fig. 7 with all the other data points removed. As shown in **Fig. 10**, the results are similar to those observed for RB-A. The contributions of high white liquor sulfidity and low sootblowing steam flow rate to fouling also stand out in this case, and they remain variables to be investigated further in this mill.

RB-C

The high resolution data available for RB-C made it possible to compare individual runs based on the number of days the boiler was operating. The “high-fouling” period comprised 47 days, from July 21 (the first day of the boiler startup after water wash) to September 5 (the day before the boiler was shut down



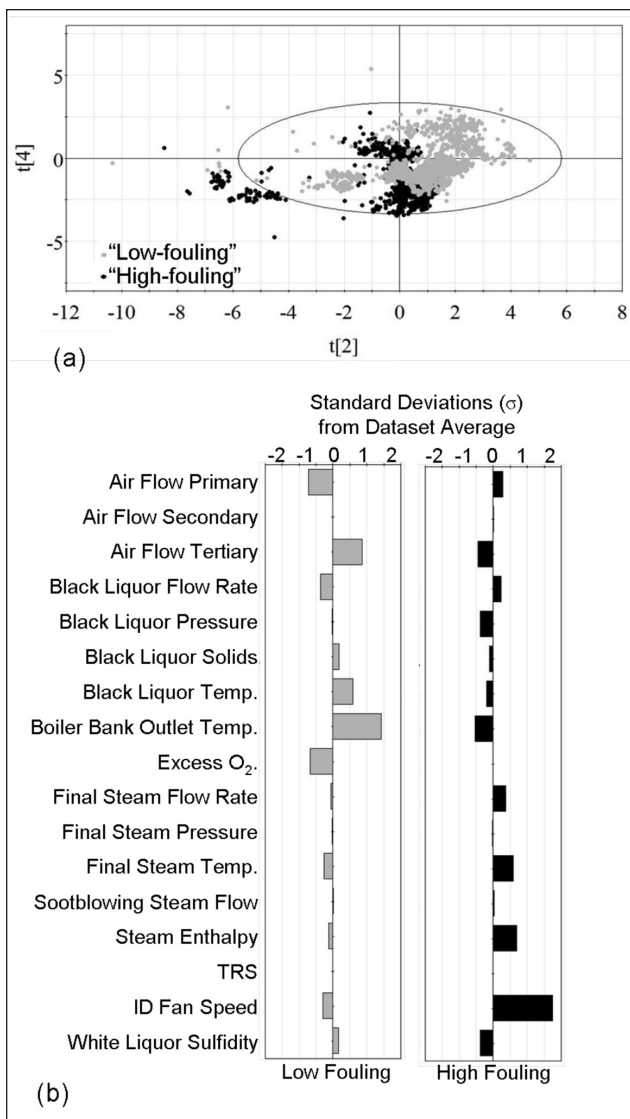
10. Comparison of low-fouling and high-fouling periods for RB-B and corresponding contribution plot.

because of plugging). The "low-fouling" period spanned 76 days, beginning July 19, and lasted 29 days longer than the high fouling period. The two periods were plotted over Fig. 8 with all the other data points removed. The two clusters of data points overlap because both periods experienced fouling, although the high-fouling period was more severe in terms of shortened boiler run (Fig. 11). Most of the data points were in the bottom-right quadrant (fouling zone) of the plot. The contribution plot (Fig. 11b) shows that mill sulfidity was low during the high fouling period of RB-C. This implies that although high sulfidity was the main factor that contributed to fouling in RB-A and RB-B, it was not the case in RB-C. The distribution of primary versus tertiary air could be a factor for RB-C, which was clearly not a factor for RB-A and RB-B.

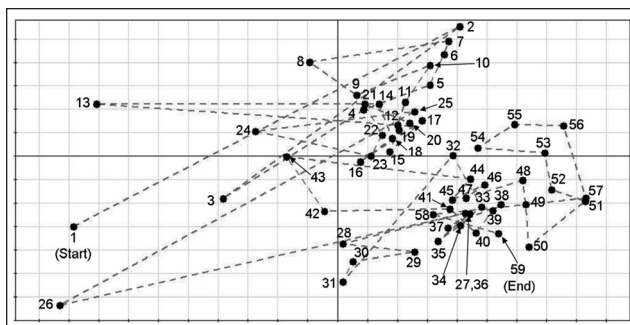
ANALYZING INDIVIDUAL RUNS

RB-A

Figure 12 shows a single run for RB-A, plotted on a magnified section of Fig. 6, with all the other points removed. The



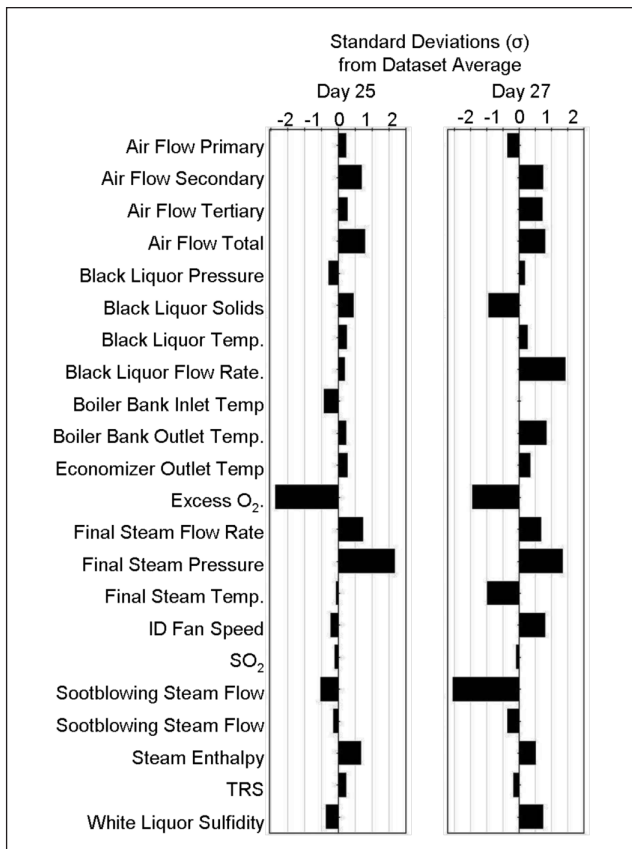
11. Comparison of low-fouling and high-fouling periods for RB-C and corresponding contribution plot.



12. A selected 59-day run for RB-A (the number of days of operation are indicated).

run was 59 days long, beginning July 28. The results show the movement of data points on the t_1/t_2 plot, beginning with the first day startup of the boiler (point 1). The boiler operated with a relatively high loading and thermal efficiency for the following 25 days (points 2-25). On day 26, the boiler oper-

RECOVERY BOILERS



13. Comparison between the day before the slowdown (Day 25 = a day of relative high thermal efficiency) and the day after the slowdown (Day 27 = a day of relative low thermal efficiency) for RB-A.

ated at a low load and low thermal efficiency, possibly resulting from a thermal shock. The thermal efficiency remained low from day 27 to day 59 when the boiler was shut down for a water wash.

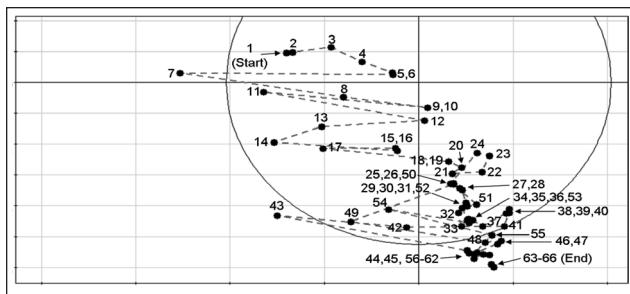
The variables during the day before the slowdown (point 25) and the day after (point 27), shown in Fig. 12, can be compared directly in a contribution plot (**Fig. 13**). Several variables stand out; the day before the slowdown had higher black liquor solids, lower liquor flow, soot-blowing steam flow and lower white liquor sulfidity than the day after. Any days (or any group of days) can be compared in similar contribution plots.

RB-B

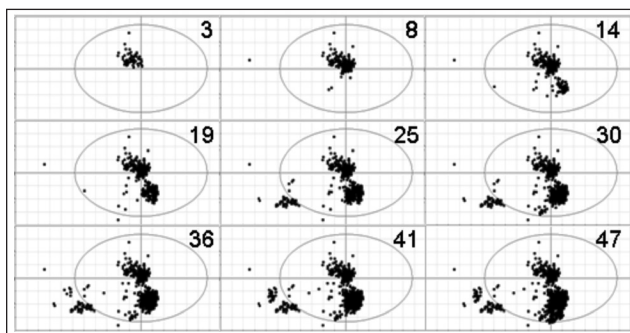
Figure 14 shows a single run for RB-B, plotted on a magnified section of Fig. 7, with all the other points removed. The run began on March 2 and lasted 66 days. A steady progression from a region of high thermal efficiency to a period of low thermal efficiency in the boiler is observed.

RB-C

Figure 15 shows a single run for RB-C, plotted on a magnified section of Fig. 8, with all the other points removed. The run is 47 days long, beginning July 21 and ending on Septem-



14. A selected 66-day run for RB-B (the number of days of operation are indicated).



15. A selected 47-day run for RB-C (the number of days of operation are indicated).

ber 5 with a water wash. Because hourly average data were used in this case, resulting in numerous data points, the plot spreads over nine frames. The plot shows a steady progression from a region of high thermal efficiency to a period of low thermal efficiency in the boiler.

CONCLUSIONS

This study used the PCA feature of a multivariate statistical program to examine potential correlations between boiler operating variables and fouling of three recovery boilers at two kraft pulp mills. The results show that for all three boilers, PCA can visually distinguish the variability related to boiler firing load and fouling, and that variables that are correlated with fouling are boiler specific. For the two recovery boilers at the same mill, high black liquor solids, high soot-blowing steam flow, and low white liquor sulfidity were correlated with improved boiler operation. For the boiler at the other mill, high primary air flow, low black liquor temperature, high black liquor pressure, and high soot-blowing steam flow were consistent with improved recovery boiler operation.

PCA also was found to be useful in visualizing gradual and sudden shifts in boiler thermal efficiency for individual runs, allowing day-to-day analysis of changes in boiler fouling conditions. This represents a major step forward in identifying operating variables that may be adjusted to minimize fouling, and in developing an on-line fouling monitoring technology based on PCA. **TJ**

Received: September 18, 2008

Accepted: August 17, 2009

ACKNOWLEDGEMENTS

This work was conducted as part of the research program on "Increasing energy and chemical recovery efficiency in the kraft process," jointly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and a consortium of the following companies: AbitibiBowater, Alstom Power, Andritz, Aracruz Celulose, Babcock & Wilcox, Boise Paper Solutions, Carter Holt Harvey, Celulose Nipo-Brasileira, Clyde-Bergemann, Diamond Power International, Domtar, DMI Peace River Pulp, Georgia Pacific, International Paper, Irving Pulp & Paper, Metso Power, MeadWestvaco, Stora Enso Research, Tembec, and Votorantim Celulose e Papel.

LITERATURE CITED

1. Tran, H.N. in *Kraft Recovery Boilers* (T.N. Adams, Ed.), TAPPI PRESS, Atlanta, Georgia, USA, 1997, Chap. 9.
2. Hodouin, D., MacGregor, J.F., Hou, M., et al., *CIM Bulletin*, 86(975): 23(1993).
3. Simoglou A., Martin E.B., and Morris A.J., *Control Engineering Practice*, 8(8): 893(2000).
4. Yin, K.K., Yang, H., and Cramer, F., *Chemical Engineering Communications*, 189(9): 1242(2002).
5. D'Souza, T., Repka, J., and Tran, H.N., "Effects of process variability on ring formation in lime kilns," *Proceedings of the 2006 TAPPI Engineering, Pulping, and Environmental Conference*, Atlanta, TAPPI PRESS.
6. Versteeg P.L., and Tran H.N. "A multivariate analysis of recovery boiler fouling," *Proceedings of the 2007 International Chemical Recovery Conference*, pp. 191-198, Pulp & Paper Canada (In Press).
7. MacGregor, J.F., and Kourti, T., *Control Engineering Practice*, 3(3): 403(1995).
8. Kourti, T., and MacGregor, J.F., *Chemometrics and Intelligent Laboratory Systems*, 28(1): 3(1995).
9. Umetrics AB, *SIMCA-P and SIMCA-P+, version 10, user guide*, Umetrics AB, Umeå, Sweden, 2002.

INSIGHTS FROM THE AUTHORS

We choose this topic for research because recovery boiler fouling and plugging can greatly hinder the production capacity of kraft pulp mills. The study reported here complements our other research projects on the effects of fireside deposits on recovery boiler fouling and corrosion.

This study showed that PCA can be used to monitor fouling in recovery boilers. Mills might use this information to devise appropriate strategies for minimizing fouling and plugging in their recovery boilers.

The most difficult aspect of this research was to obtain sufficient operating data so that the PCA results obtained are meaningful. We addressed it by discussing the issues with recovery staff at various mills and by performing tests using data sets with

different time intervals.

The next step in this research is to work with interested mills to develop a fouling monitoring system based on PCA.



Versteeg



Tran

Versteeg and Tran are with the Pulp & Paper Centre and Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada; email Tran at honghi.tran@utoronto.ca.