

Leveraging mill-wide big data sets for process and quality improvement in paperboard production

JIANZHONG FU AND PETER W. HART

ABSTRACT: The MWV mill in Covington, VA, USA, experienced a long term trend of increasing episodes of paper indents that resulted in significant quantities of internal rejects and production downtime. When traditional troubleshooting techniques failed to resolve the problem, big data analysis techniques were employed to help determine root causes of this negative and increasingly frequent situation. Nearly 6000 operating variables were selected for a deep dive, multi-year analysis after reviewing mill-wide process logs and 60000+ PI tags (data points) collected from one of the major data historian systems at the MWV Covington mill. Nine billion data points were collected from November 2011 to August 2014. Strategies and methods were developed to format, clean, classify, and sort the various data sets to compensate for process lag time and to align timestamps, as well as to rank potential causes or indicators. GE Intelligent Platforms software was employed to develop decision trees for root cause analysis. Insights and possible correlations that were previously invisible or ignored were obtained across the mill, from pulping, bleaching, and chemical recovery to the papermaking process. Several findings led the mill to revise selected process targets and to reconsider a step change in the drying process. These changes have exhibited significant impacts on the mill's product quality, cost, and market performance. Mill-wide communications of the identified results helped transform the findings into executable actions. Several projects were initiated.

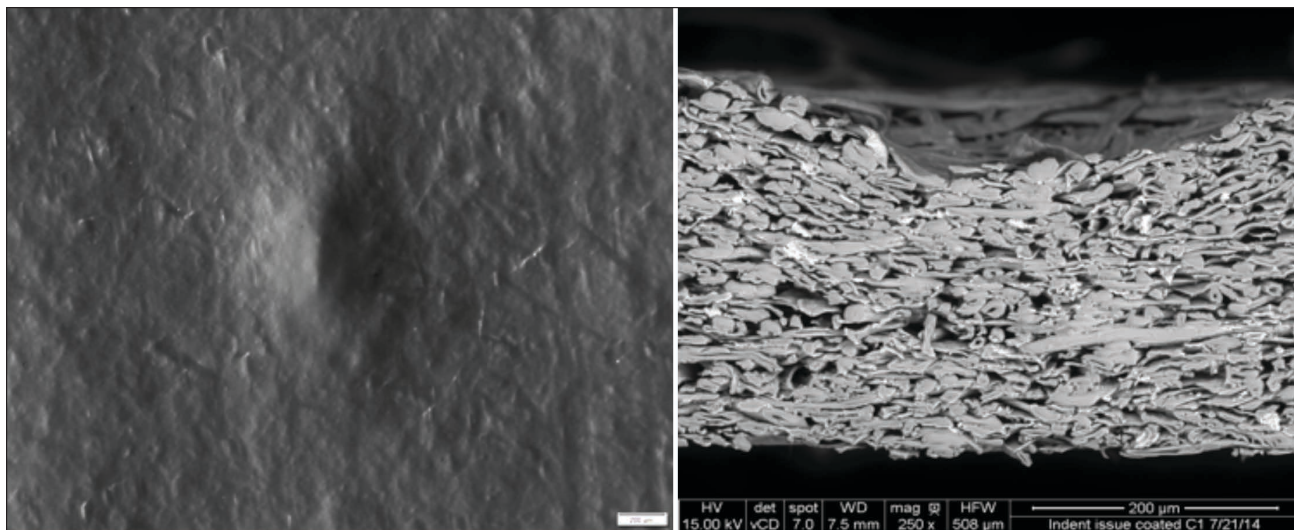
Application: Papermaking, like other manufacturing, is facing an unprecedented explosion in data collection and availability. Big data analytics can develop predictive models and find critical insights that are essential to improve operational performance.

In the papermaking industry, indents are common paperboard defects that cause white specks in printed images. Poor print quality jeopardizes brand recognition and results in complaints, rejected product, and costly restoration of customer relationships. Indents can be described as surface deformations resulting from built-up paper debris or various deposits that are pressed against the paper web. Indents could also be caused by melt away components such as acidified rosin soap (waxes), oils, or other low melting point hydrophobic materials from broke that accumulate on the sheet surface and later melt away during the drying process, leaving behind a void or indent [1]. Indentations occurring during the papermaking process typically involve fibrous and coating materials that stick to one or more dryer cans, felts, and other nip points. The consequence of a significant amount of indents is often unplanned shutdowns to clean up the machine and physically remove the deposits. An internal evaluation performed in early 2014 estimated that USD 2.8 million annual savings could be realized by eliminating indents and downtime due to wash-ups on just one paper machine at the MWV mill in Covington, VA, USA. **Figure 1** shows typical indents from the Covington mill's No. 1 paper machine (PM No. 1) from both a top view and in cross section to show the depth of the indent.

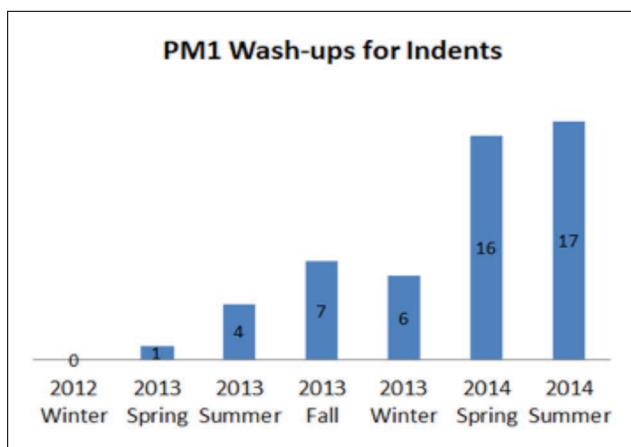
Since fall of 2013, the Covington mill's PM No. 1 has been experiencing an unusual number of indentation issues — nearly 10 times higher than historical values. **Figure 2** shows the results of these increased indentation issues as a substantial increase in the frequency of wash-ups. Wash-ups occur when the internal reject rate increases to the point where operators determine it is more economical to stop production and physically clean buildup and debris off the drier cans. In order to accomplish the cleanup, the paper web has to be broken back to the wet end of the paper machine, which results in lost production. The increase in wash-ups shown in Fig. 1 resulted in up to a 29% net tonnage decrease in production on PM No. 1.

For more than a year, PM No. 1 operators fought various indentation issues in the cross machine direction (CD). The indentation problem was only solved for short periods of time. These issues continued to recur, which indicated the root cause of the problem had not been properly identified and corrected. Operational troubleshooting procedures that worked for a specific set of indentation issues did not appear to work for new indentation outbreaks.

A task force, the C1 indents data team, was formed in early 2014 and a Kepner-Tregoe problem analysis was conducted [2]. The team identified that indentation issues mostly



1. Micrographs of a typical indent obtained from the No. 1 paper machine (PM No. 1) — top view (left) and cross section.



2. Number of wash-ups on PM No. 1 resulting from indent events.

occurred at the main dryer section due to buildup of material on the wire side (WS) dryer cans and were neither specific to one grade or one caliper. The main components of the material consisted of fibers and starch.

A hypothesis was made that PM No. 1 had a more complex wet end operating window that included increased sheet moisture going into the first dryer section. In addition to high sheet moisture entering the first dryer section, fiber bonding strength, equipment conditions, and operational parameters — such as dryer can doctors, refiner plates, steam pressure control, broke and white water treatment, etc. — could also strongly impact the onset of indents. These variables are wide ranging across the entire integrated mill and influenced by operations from pulping and bleaching to stock preparation, wet end chemistry, and the forming and drying processes. In order to troubleshoot and assign a root cause to the indentation issues, it was decided to look at operational data across all of these areas.

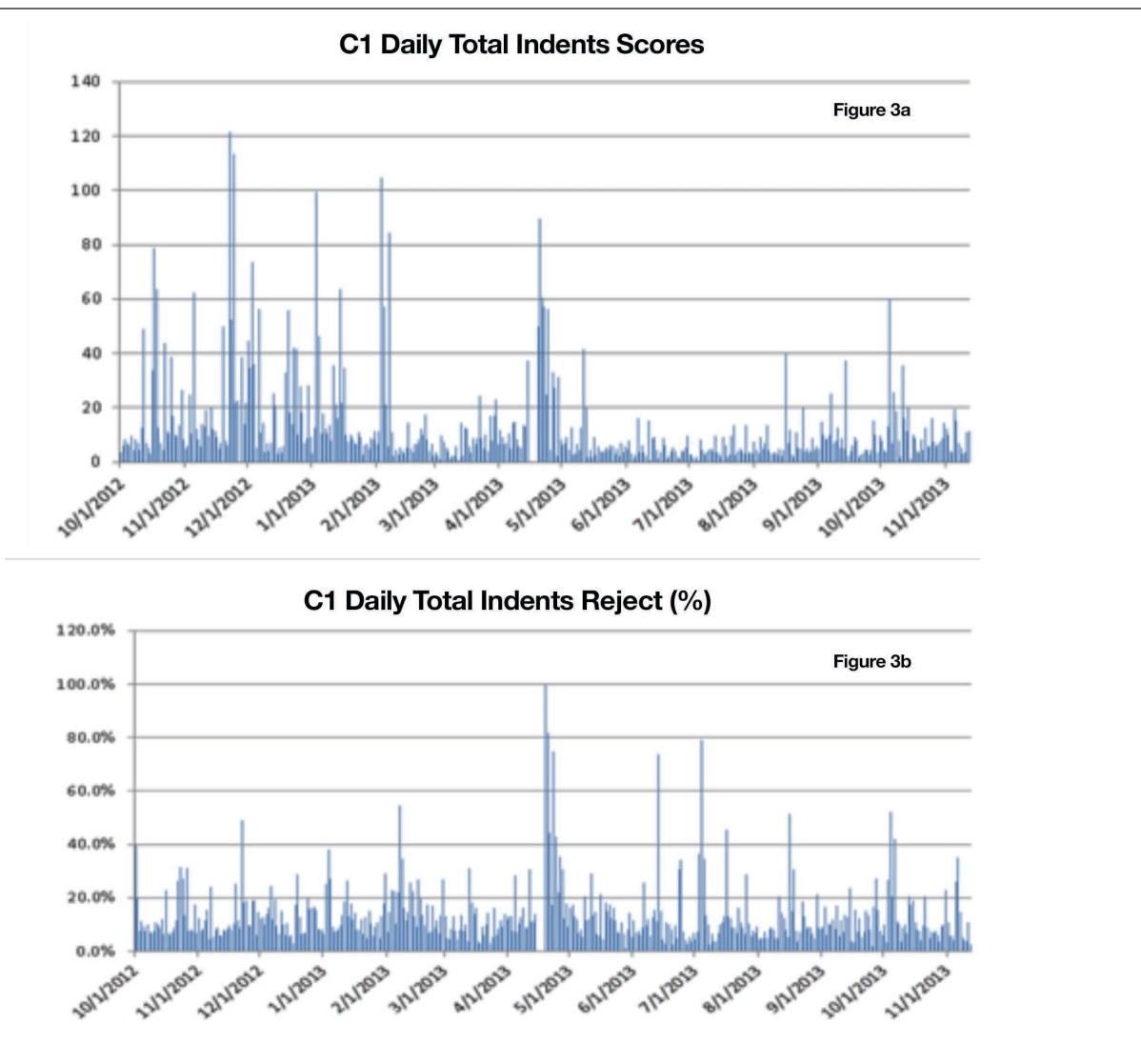
Operational data are stored in multiple formats across the entire mill. Currently, major data systems in the Covington

mill include Meridium, SAP, Metso, Optivision/Cold Fusion, PI, and Proficy, in addition to manual entry area operation logs and databases. To troubleshoot root causes by utilizing such complicated and disparate “big data” systems is a daunting and difficult task. In particular, PI is a data historian that tracks and saves data for long term archival purposes, automatically entering data points at a routine time from 1 per second to one every 30 min, depending on how the data point is tagged into the system. One of the PI servers in the Covington mill has more than 60000 tags (operational data points of some type). This PI server records data in non-uniform rates, such as every 12 s and once a shift. The resulting amount of data is enormous. For instance, if the PI server recorded data once per minute, a three year analysis of a single process variable would require the analysis of more than 1.5 million data points. Therefore, a significant part of the indents data team’s assignment was to collect, process, and analyze mill-wide big data sets. The GE Intelligent Platforms (CSense Troubleshooter) program had been successfully employed by the mill in another big data analysis to identify potential areas of improvement for a quality-related issue, and this program was employed as a potential tool in the current analysis.

DATA COLLECTION AND PROCESSING

The indents data team initially used daily total indent scores calculated from the Metso Web Inspection system, and daily total indents rejects (%) as targets to identify periods of high versus low indents. When compared over a three-year time period, these two data sets did not correlate. **Figures 3** shows the daily total indent scores from the web inspection system (Fig. 3a) and the percent daily total rejects (Fig. 3b) for the same timeframe. A comparison of the two graphs shows little useful correlation.

As a result of the poor correlation, the Metso data were excluded from our analysis. It is possible the inspection system is mistakenly categorizing other defects as indents. It is



3. Daily total indent scores obtained from the Metso web inspection system (3a) and the percent daily rejects for indents (3b). Both graphs are for the same timeframe.

clearly shown in the first half of Fig. 3 that there were no significant indent rejects, but the web inspection data showed significant levels of indents. The Metso defects recognition system is currently being improved to better recognize and categorize indentation events.

Since the indentation issues occurred mostly at the main dryer section and were not grade or caliper specific, data were analyzed on a daily basis instead of diving into details by grade, by caliper, or by reel. The indents data team also decided to exclude all operational variables after the paper machine size press and focus on pulping, bleaching, coating preparation, and paper machine operations.

Process data used for analysis were collected from one of the major PI servers and manual entry area operations logs. Other data sources were mainly used for filters and verification. The team first divided into four groups (pulping, starch kitchen, paper machine, and quality control) to subjectively

exclude irrelevant variables, variables with all zero values, and obviously redundant data tags. The driving mantra for excluding any data point was, “If there is any possible or conceivable way this process variable could impact pulp or paper quality, keep it in the initial analysis.” The team dug into process logs to obtain the status of physical process changes, such as roll changes, felt types, number of days a specific felt had been used, and specific operational target changes in the cooking and bleaching processes. This subjective, but expert knowledge-driven data filtering process reduced the number of variables used in the analysis from nearly 60000 to below 8000.

The indents data team decided that the PI data would be sampled at a rate of one data point every minute for all numeric variables (continuous and discrete). Daily means and variances would be determined for each set of 1440 data points collected. It was also decided that for all class or string variables, the data would be digitized at a scale from 1 to 1440;

PAPERMAKING

e.g., if a valve was “ON” for an hour today, then its value would be 60 for today. Similarly, if the valve was “ON” for 32 min, its value would be 32 for that day. Typical patterns of digitized class variables are illustrated in **Figure 4**.

Once the specific PI data points to be used in the data analysis were identified, fifteen desktop personal computers were programed to pull PI data for each variable. A timeframe of November 2011 to August 2014 was set as the boundaries for the data pull. About one week was required to complete this task. Another two weeks were required to extract variables from process logs and manually enter them into usable form (Microsoft Excel spreadsheets) for the analysis. In total, about 8000 data sets were obtained from the timeframe of interest. These data sets were further reduced to 6000 variables by sorting the daily mean and variance, along with manually excluding irrelevant class variables.

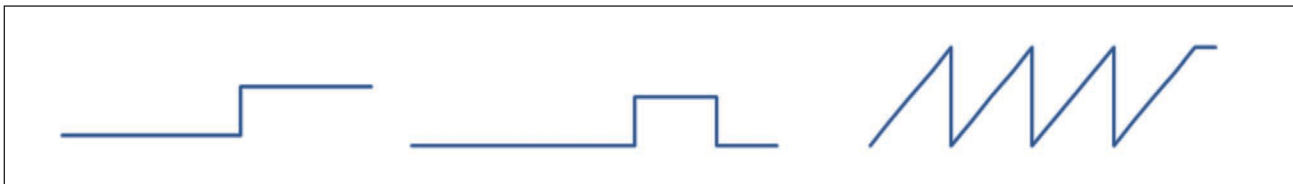
DATA ANALYSIS

The data sets obtained from PI and operational logs were analyzed using the GE Intelligent Platforms (CSense Troubleshooter) computer program [3]. Decision trees were prepared by the program to rank potential causations for the indentation issue. Decision trees are rule-based process models. A typical decision tree model is shown in **Fig. 5** [4-6]. The CSense decision tree program applies three distinct algorithms (entropy

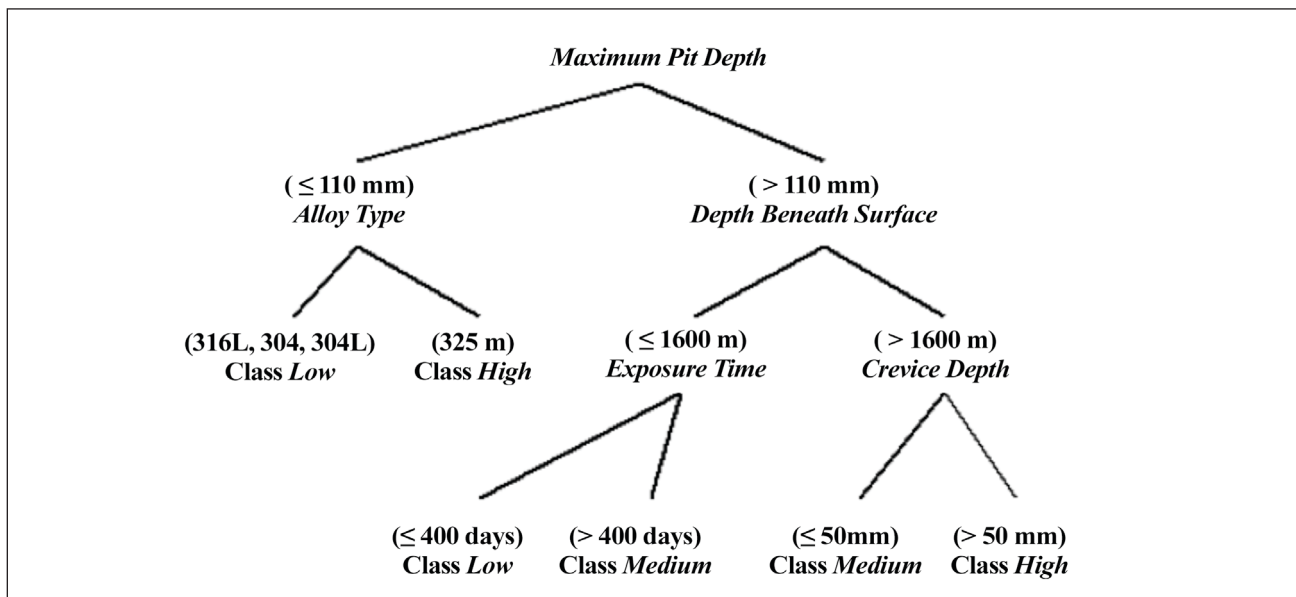
gain, chi square, and Gini index) to provide the maximum information in one class of process variable. The algorithms divide the data set into subpopulations with significantly different distributions and a single group of as large a size as possible [4,5]. The decision tree identifies the best way to explain the dependent variables as a function of the independent variables.

The indents data team explored several paths and decided to use the following strategies for troubleshooting the indentation problem:

1. Determine the indentation baseline time period and high reject time period.
2. Define the indents event days and non-event days:
 - Indents event days had to meet all of the following criteria:
 - Paper machine shutdown occurred to clean up the indents.
 - The indents were found to originate in the main dryer section.
 - Daily total rejects for indents were greater than 3% of the gross production.
 - Daily gross production of the paper machine was greater than 500 tons.
 - Non-event days had to meet the following criteria:
 - Zero downtime.



4. Typical patterns for digitized class variables.



5. Typical decision tree model resulting from a CSense decision tree analysis employing the GE Intelligent Platforms (CSense Troubleshooter) computer program.

- No rejects for indents.
 - Daily gross production of the paper machine was greater than 800 tons.
3. Generalize paper machine lag time to other operational areas mill wide so that the timestamps of potential causes and an indent event are approximately matched.
 4. Develop decision trees by operational areas to identify the top 15-50 variables and then combine selected variables for final analysis.

RESULTS AND DATA VALIDATION

For the analysis, November 2011 to August 2013 was used as the baseline time period and September 2013 to August 2014 as the high indents reject time period. Fifteen event days were identified within the high indents reject time period, and 230 non-event days were identified within the entire time range (both baseline and high indents reject periods). The indents data team established lag times for operational areas based upon in-depth process knowledge, i.e., applying an 18 h delay for all variables from the pulp mill and an 8 h delay for all variables from the bleach plant. Potential causations and indicators of indentations were determined from the top ranked variables obtained from the decision trees developed using the GE Intelligence Platforms program. These potential causations, as highlighted in the following sections, were obtained by comparing indents event days with non-event days from the high reject and baseline time periods.

Event vs. non-event days during high reject time period — potential causes

- Unbleached pulp mill: washing performance, hardwood and pine shives, and white liquor chemistry.
- Bleach plant: brown water filtrate level, shower water, brightness, and bleach chemicals.
- Beater room: pine pulp pH, broke transfer, ultraphase size usage, consistency control and dilution water, non B-Unit pulp (pulp not from the B bleach plant), and stock split ratio between B and C bleach plants.
- Coating and additives: coating and size press microbiological counts, biocide residue, starch, defoamer, and viscosity modifier.
- Paper machine: effluent flow, stuffbox pH, white water chest level, rectifier roll speed, lumpbreaker roll nip pressure, wire turn roll load, blue dye flow, press load and vacuum, main dryer section (steam pressure, flash flow control, and steam superheat temperature control), and moisture variation.

Event days of high reject time period vs. non-event days of baseline time period — potential causes

- Unbleached pulp mill: pine and hardwood knotters, shives, digester additives, brownstock washer carryover, white liquor chemistry, and kappa number.
- Bleach plant: white water flow from paper machine, brownstock washer carryover, and washing performance.

- Beater room: consistency control and dilution water, broke transfer, non B-Unit pulp, and pine chest level.
- Coating and additives: coating and size press microbiological counts, dispersant, and viscosity modifier.
- Paper machine: water system (cloudy chest level, saveall, white water, effluent flow), wet end chemistry (alum addition, dye addition), stock system (pine chest level, hardwood refiner, cleaner and screen rejects, broke system), headbox (differential pressure, shower water temperature, rectifier roll), vacuum system (couch, duoformer, flatbox vacuum, vacuum pump seal water), drying system (steam temperature, condensate, base sheet moisture), and CD control (raw stock basis weight, caliper).

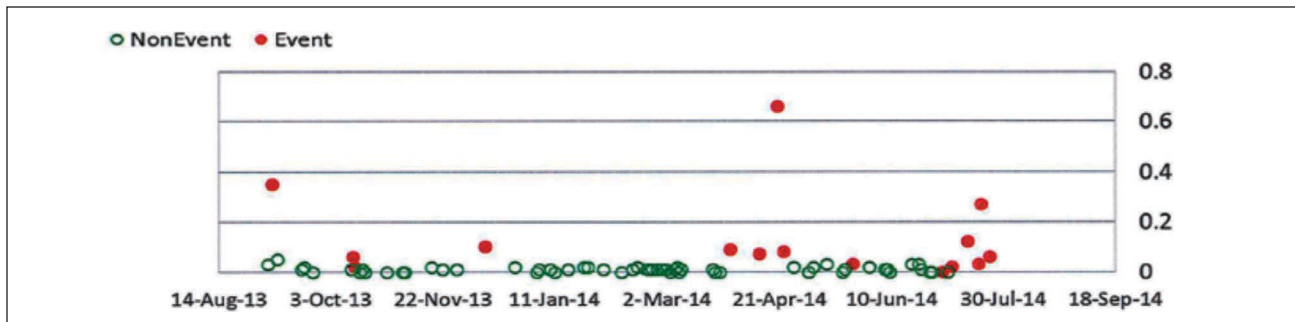
The indents data team then carefully evaluated the data for the potential causation variables and reviewed the findings with operational area experts in pulping/bleaching, coating, and papermaking to validate the findings. Several of the variables among the top ranked data sets were excluded due to bad data that had not been eliminated in the first round of data filtering. Some of these bad data points did not exclusively exclude paper machine downtime in the indents event day timeframe. Downtime was found to dramatically skew several of the results and artificially place the specific data points into potential causation slots in the decision trees.

Time-based scatter plots of the data were prepared for the indents event and non-event days and cross-referenced variable descriptions, along with process locations, to investigate the findings obtained from the decision trees. The combined results from the previous two comparisons within baseline and high reject time periods are discussed in the following.

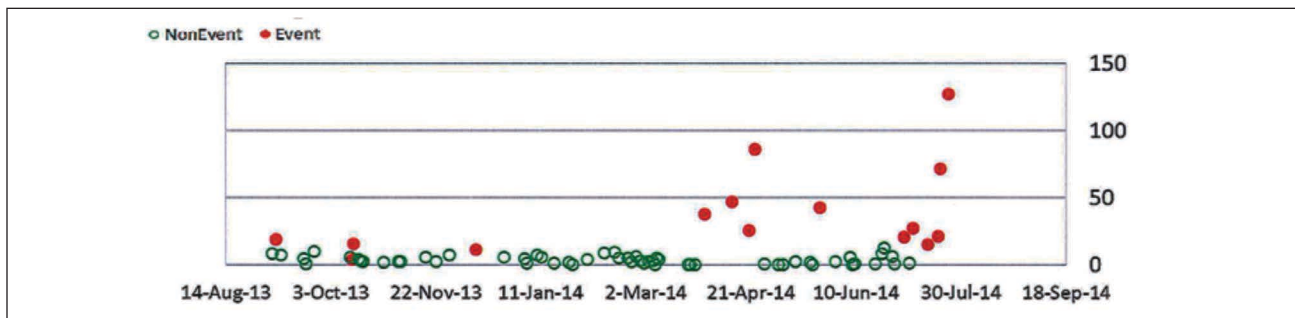
In the pulp mill, the high variances of washer liquor baume, shive counts, white liquor chemical strength (total titratable alkali [TTA], sulfidity, and active alkali), and brown water loop valve open positions were found to be associated with the indent event days. The Covington mill uses a semi-closed water loop to dilute washed brownstock pulp in high density storage tanks and transports it to deckers feeding the bleach plant feed tanks; when there is not enough evaporator condensate for brownstock washing, water is taken from this water loop. The valve position represents paper mill white water makeup into the water loop.

As shown in the scatter plot figures that follow, the red dots represent indents event days and the green circles represent non-event days. **Figures 6-8** illustrate that most of the event days occurred when these variances were high.

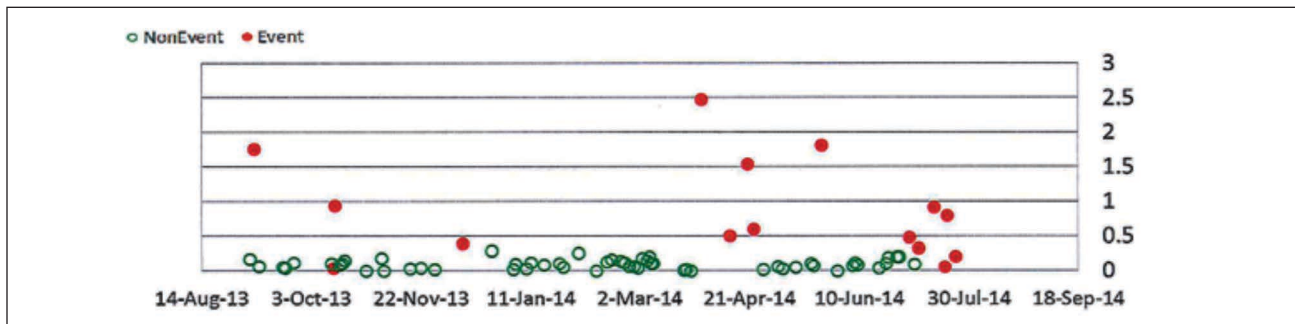
The bleached pulp mill consists of three distinct bleaching units. A Unit bleaches oxygen delignified pine in a four stage D(EP)DD bleaching sequence. B Unit bleaches non-oxygen delignified hardwood brownstock in a four stage D(EP)DD bleaching sequence. C Unit bleaches oxygen delignified hardwood in a three stage D(EOP)D bleaching sequence. In the bleached pulp mill, the amount of wash water used, pulp brightness, and A-Unit (pine) bleach chemicals were found to be associated with the indent event days. **Figures 9-10**



6. Variance of average liquor baume. Ten out of 11 days were indents event days when the variance was higher than 0.04.



7. Variance of shive counts. Fourteen out of 15 days were indents event days when the variance was higher than 11.



8. Variance of average active alkali. Twelve out of 14 days were indents event days when the variance was higher than 0.3.

illustrate that significant indents event days correspond to lower gallons per ton of the shower water usage and higher brightness. **Figures 11-12** show the majority of indents event days correspond with low variance of chlorine dioxide (ClO_2) flow and high ClO_2 flow.

In the beater room, the KABOK-to-No. 4 high density storage tank (HD), representing the bleached pine flow to the No. 4 HD; purchased pine pulp usage; No. 12 chest pH; and No. 4 HD consistency water flow (to dilute pine consistency) were associated with high indents event days. The split ratio to PM No. 1 from B/C-Units (C Unit bleaches oxygen delignified pulp in three stages, while B Unit bleaches non-oxygen delignified pulp in four bleaching stages) was also associated with high indents event days. **Figures 13-14** show that most indents event days correspond with high pine flow from A Unit and low usage of purchased pine. **Figure 15** illustrates that significant indents event days correspond to lower split ratios to PM No. 1 from B/C-Units.

At the paper machine, broke usage, blue dye flow, rectifier

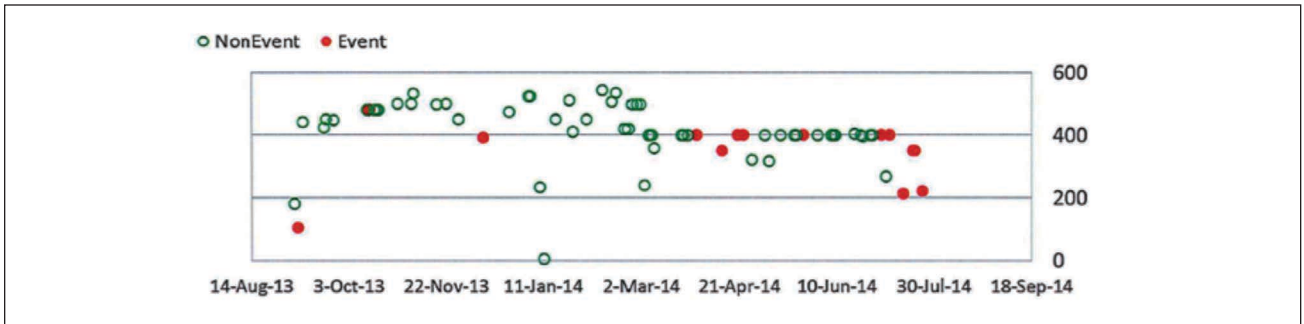
roll speed, wire turning roll load, lumpbreaker roll load, first and second press load and vacuum, vacuum pump seal water pressure, and first dryer section steam pressure were associated with event days. **Figures 16-18** show the majority of indents event days correspond with the low press loads, reduced vacuum, and low steam pressure.

Sheet properties also showed some correlation with indents event days. Significant indents event days were associated with uneven raw stock, CD moisture profile, average caliper, moisture 2-sigma variation, and raw stock weight. **Figure 19** illustrates that high degree of variation in sheet moisture corresponds to a high frequency of indents event days.

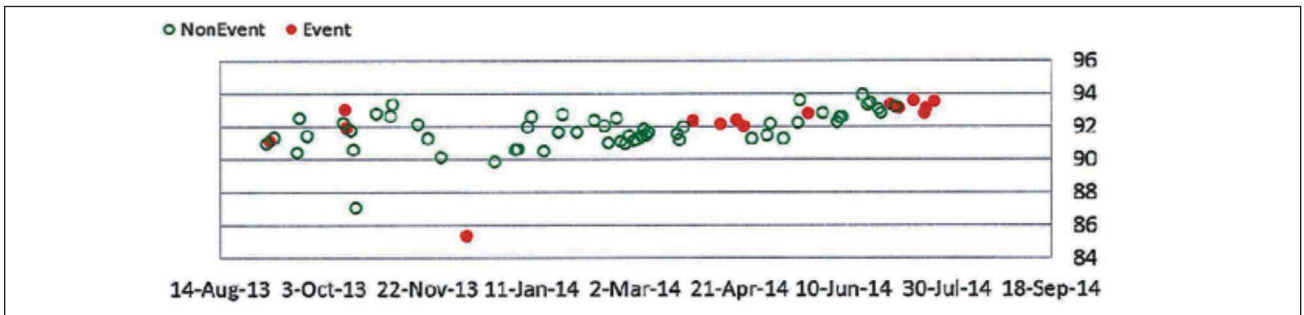
KEY INSIGHTS

Fiber strength

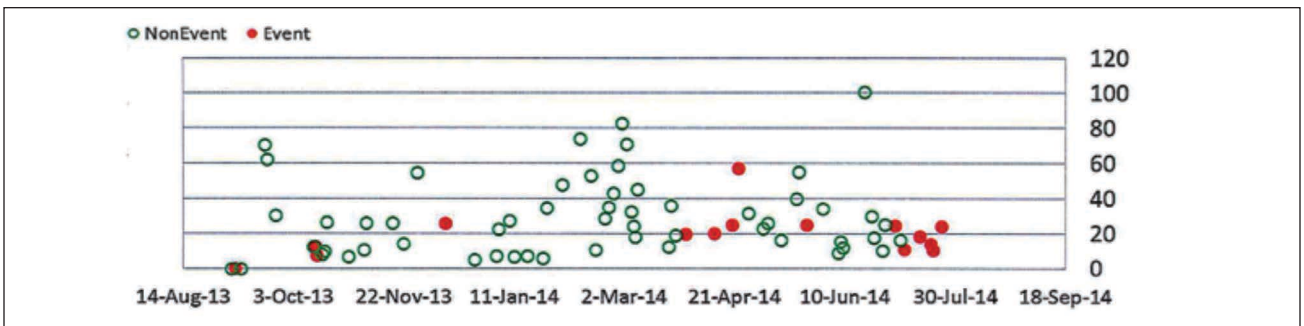
Several of the key variables in the pulp mill and bleach plant areas suggested operational methods of producing pulp with reduced strength or reduced bonding potential. For instance,



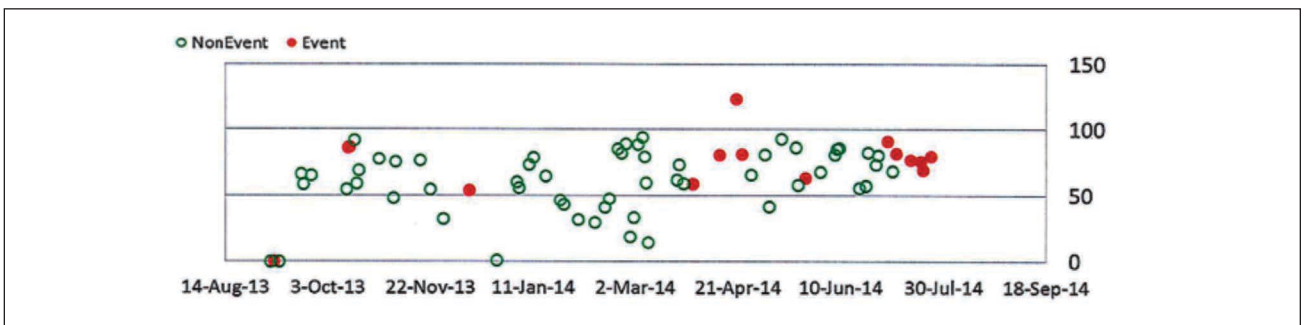
9. Mean of shower water usage. Two out of 31 days were indents event days when the mean was higher than 406 gpm, and 13 of the 15 event days are associated with lower shower water usage.



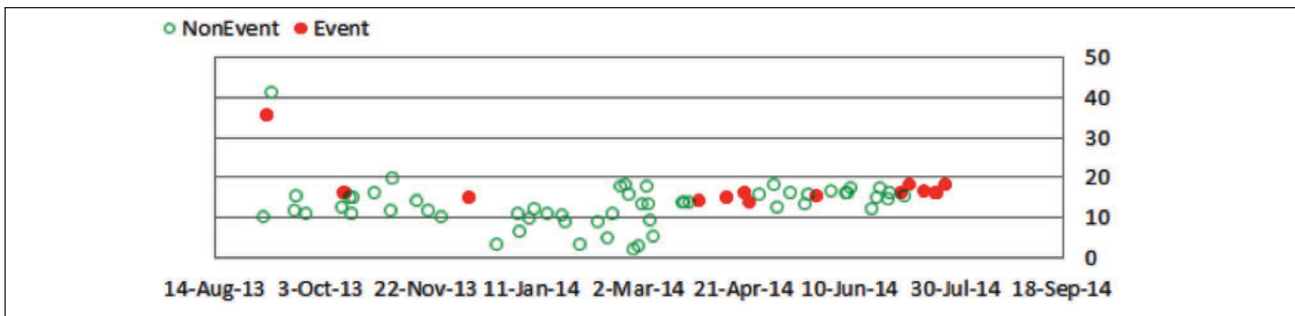
10. Mean of brightness to bleached. Two out of 32 days were indents event days when the mean was lower than 92, and 13 of the 15 event days correspond to the higher brightness.



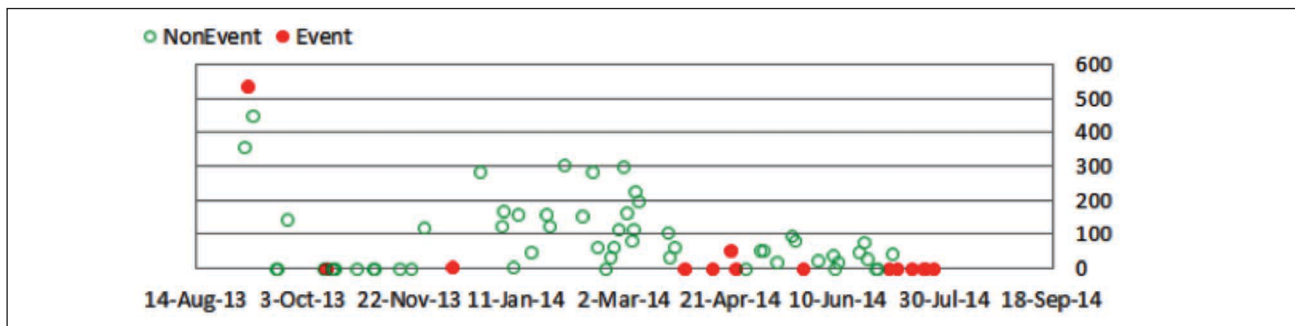
11. Chlorine dioxide (ClO_2) flow (gal/min). Majority of indents event days occurred when the variance of ClO_2 was low.



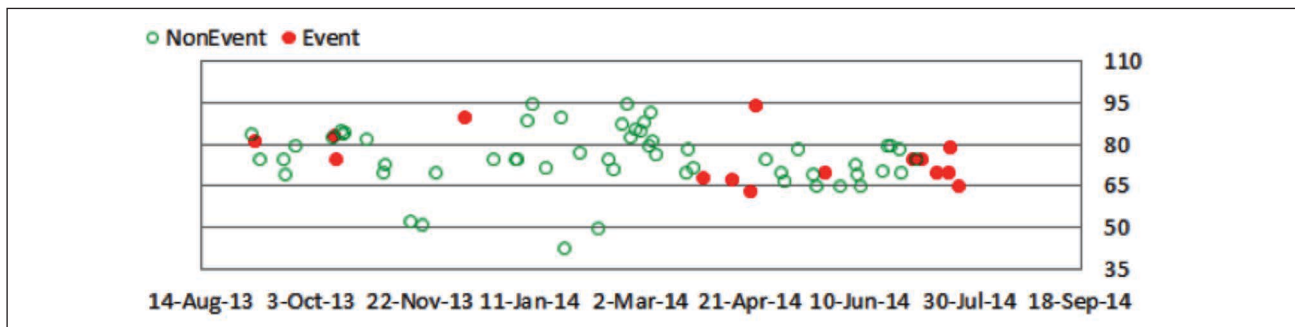
12. Chlorine dioxide (ClO_2) flow (gal/min). Majority of indents event days occurred when the mean of ClO_2 was high.



13. Mean of A-Unit production rate (tons/h). Fourteen of 15 indents event days are associated with higher pine production rate.



14 Mean of purchased pine usage (ton). Fourteen of the 15 indents event days are associated with lower usage of purchased pine.



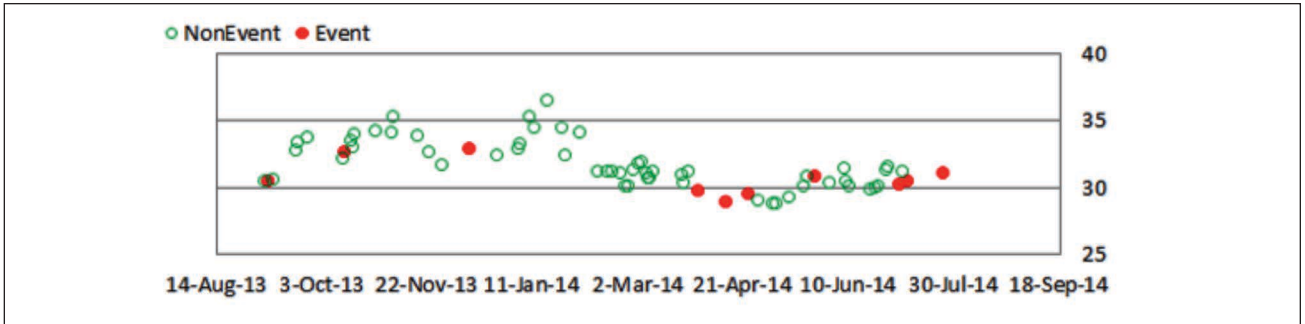
15. Mean of the split ratio to PM No. 1 from B-/C-Units (%). Majority of indents event days are associated with lower split ratio to PM No. 1 from B-/C-Units.

when brownstock soda carryover from the unbleached pulp mill to the non-oxygen treated bleaching line is low, indents event days are limited, potentially corresponding to easier-to-bleach pulp; when the variability of washing is elevated, the number of event days are high, potentially resulting from over-bleaching to catch the higher carryover periods. When the amount of shives going to the bleach plant is changing substantially, the tendency of the bleach plant operator is to over-bleach to reduce dirt and pulp contamination. It is possible that over-bleaching for any of these reasons could result in a loss of pulp strength or bonding capability. The weaker pulp would then have a greater tendency to pick from the sheet onto the first set of dryer drums, collect into deposits, and result in indents.

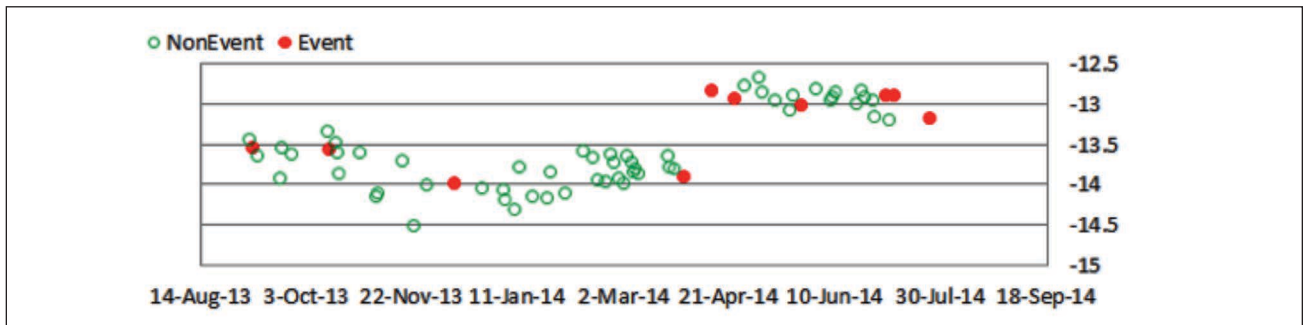
When the white liquor strength variability is high, the mill experienced indents event days. Uniformity of white liquor strength is important to maintain stability in kappa number

and pulp quality. Again, variability in kappa numbers could also result in over-bleached, weaker fiber being sent to the paper machines. Days with low wash water per ton of pulp ratio correspond to higher carryover between stages and have a potential to negatively impact pulp strength. Finally, higher brightness indicates over-bleached pulp, which could potentially hurt strength.

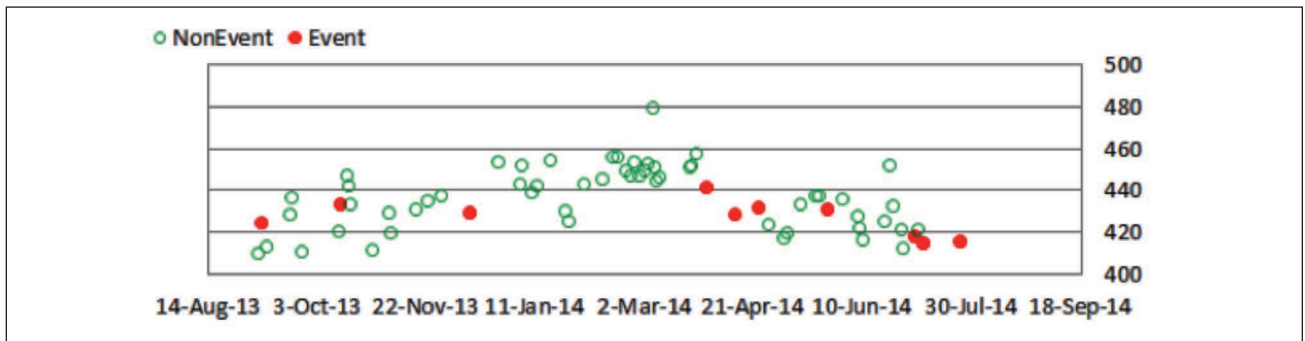
When all of these potential causation values were closely examined, the variance in ClO_2 flow being low was the determining factor that rated the highest. The majority of low variability days corresponded with very high ClO_2 flow when compared with the mean for the same. The combination of high ClO_2 flow, low variance, and high event days strongly suggests that the bleach plant is not reacting fast enough to changes in: (1) incoming kappa number, and (2) brownstock washer carryover, which results in production of fiber that is over-bleached and has lower strength and reduced bonding potential.



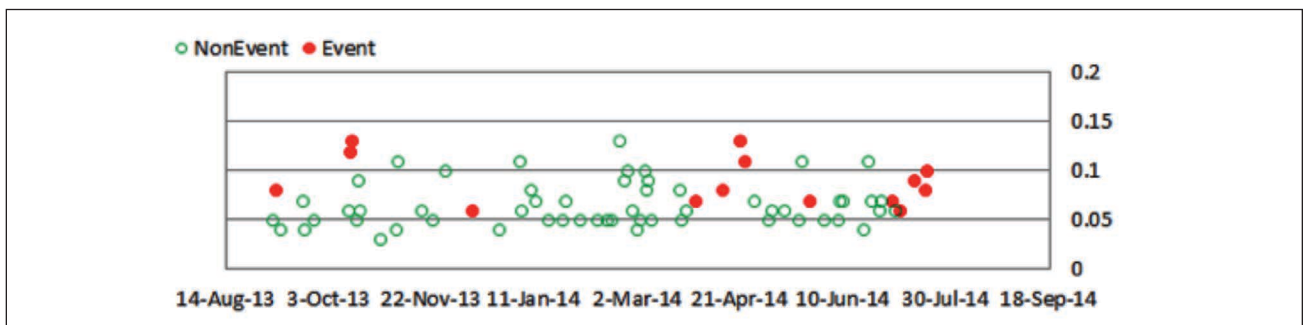
16. Mean of smooth press bottom loads (%). Majority of indents event days are associated with lower press loads.



17. Mean of first press vacuum (Hg). Majority of indents event days are associated with lower press vacuums.



18. Mean of first dryer section steam temperature (°F). Most of the indents event days are associated with lower steam temperature.



19. Mean of reel moisture variation 2-sigma machine direction long term variation (%). Majority of indents event days are associated with high moisture variation.

PAPERMAKING

Refiner operations could potentially be modified to improve physical strength properties during times of lower strength pulp production. Unfortunately, in order to adopt this approach, lower pulp strength conditions would have to be known in advance and refining modifications would be required to counter the lower strength properties. Refining variables were included in the big data analysis, but because the refining conditions are typically set and maintained, refining variations were not identified as potential root causes of indent formation, and refiner changes were not evaluated in the current study.

Paper machine variations

The high indents event days strongly correlate to low press loads, reduced press vacuum, and low steam pressure, which suggest that the amount of water removed from the press and main dryer section are reduced. Reduced water removal results in a wetter sheet and increased tendency to pick fibers from the sheet onto the dryer drums; thus, there is a higher chance to build up debris and form indents.

According to the decision tree analysis, when the smoothing press load was lower than 32%, 8 out of 10 event days occurred. The analysis suggests that lower smoothing press loads double the chance of indent formation compared with higher press loads. Similarly, 6 out of 10 event days occurred when the first press vacuum was reduced to less than -13.5 in. of Hg. The chance of indent formation at lower press vacuums is 150% higher than that at greater negative vacuums. Finally, 9 out of 10 event days occurred when the dryer steam temperature was below 227°C; the chance for event occurrence increased 400% compared with that when the dryer temperature was above 227°C.

All of these paper machine variables result in a higher moisture content sheet leaving the wet end of the machine and entering the dryers. The wetter sheet is not as tightly bound, and fibers and starch can be picked from the sheet onto the dryer cans. Eventually, these fibrous spots build up on the dryers and start pressing into the paper web, forming indents.

Project efforts

As a result of developing the overall understanding of lower fiber strength and a wetter sheet entering the main drier section as potential causes of indents, several specific projects were initiated in the mill to minimize or reduce the identified causation events. Projects in the pulp mill and bleach plant areas included the following:

- White liquor causticizing probes were repaired and put back into service in an effort to minimize white liquor variability to the digesters.
- Improved operator response to incoming changes and carryover to the bleach plant were initiated to reduce the over-application of ClO_2 during the first stage of pulping.
- Improved communication of shive levels in the pulp were established between the unbleached pulp mill and the bleaching units.

- Improved first stage bleach plant washing was achieved to reduce carryover from the first ClO_2 stage to the extraction stage.
- Efforts were made to ensure that brightness targets were met without significant levels of over-bleaching to higher-than-required pulp brightness levels.

Potential projects were identified for the paper machine area, including:

- Increasing the press loads.
- Increasing the vacuum level on the first press.
- Increasing the drier temperature to 440°F or greater.

Several of these projects were successfully implemented. A few of them (mostly around the paper machine area) were not implemented due to the potential for unforeseen operational impacts. In addition to these specific projects, several key operating parameter set points were modified to bring them closer to the set points employed during non-event day operating conditions.

CONCLUSIONS

Daily averages and variances of about 6000 variables were successfully pulled from the mill data historian and analyzed for potential causation events related to indent formation on PM No. 1. Pulp mill and liquor cycle upsets, over-bleaching events in the bleach plants, wet end operating conditions, press loads, and dryer steam temperatures were all determined to participate in indent formation. Cooking liquor strength swings and pulp mill and bleach plant conditions that could result in reduced fiber strength and reduced fiber bonding potential were determined to be one set of root causes for indents. As discussed in the Key Insights section for the paper machine, any paper machine operations that result in a higher moisture content sheet entering the dryer section or reduced drying in the first dryer section were determined to be another major set of root causes.

As a result of the identification of these root causes, mill process settings were modified in an effort to mitigate several of these potential root causes. Since these process changes have been implemented, the occurrence of main dryer related indents has substantially decreased. Reduced paper machine downtime and reduced internal rejects have resulted from these changes. These results strongly suggest the root causes identified by the big data analysis approach in conjunction with in-depth process knowledge were successfully combined to diagnose and eliminate several of the root causes of indent formation on PM No. 1. The evaluation and use of the big data sets, as well as the interpretation of the decision tree results, required people with in-depth process knowledge to obtain successful results. **TJ**

ACKNOWLEDGEMENTS

The authors thank Cullen Alexander, Gregg Brelsford, Wayne Bryant, Christopher Brzak, Aaron Crumpler, Teresa Krug, Ste-

ven Parker, Craig Walton, and Donna Walton from MWV for their contributions and insights to this work. Nicholas Kern and Al Mildenerger of the Clear Group LLC were also integral parts of the C1 indents data team.

LITERATURE CITED

1. Iggesund, *Iggesund Paperboard Reference Manual*, "Unwanted surface defects," p. 95. Available [Online] http://www.iggesund.com/Global/Iggesund%20documents/RM%20Pdf/2.%20Surface%20and%20appearance/Unwanted_surface_defects_EN.pdf <9May2016>.
2. Anonymous, "Kepner-Tregoe Problem Analysis," *Quality Matters*, 14 January 2013. Available [Online] <https://iancos.wordpress.com/2013/01/14/kepner-tregoe-problem-analysis/> <9May2016>.
3. General Electric, *GE Proficy C Sense Manual*. Available [Online] http://www.vertice.com.uy/Download/Proficy_CSense.pdf <9May2016>.
4. General Electric, *GE Proficy C Sense*, program help file.
5. Berson, Alex, Smith, Stephen, and Thearling, Kurt, "An Overview of Data Mining Techniques," from *Building Data Mining Applications for CRM*, McGraw-Hill, 2000. Available [Online] <http://www.stat.ucla.edu/~hqxu/stat19/DM-Techniques.pdf> <9May2016>.
6. Kaushal, A. and Shukla, M., *Int. J. Comput. Sci. Inf. Technol.* 5(1): 651(2014).

ABOUT THE AUTHORS

We chose to research this topic because traditional troubleshooting techniques were not getting to the root cause of indent formation on the paper machine. As a result, alternative methods had to be employed to fix the problem. Application of big data analysis techniques to this issue provides another tool for mills to use when troubleshooting persistent problems that traditional methods cannot easily solve.

A significant amount of effort was required to clean up and analyze the data. Once the data were successfully cleaned, grouped, and sorted, we had to determine which sets could be potentially viable. Several group meetings were held with operators and industry experts in an effort to successfully eliminate events that correlated with problems but were not direct contributors to the problem.

An interesting finding from this research was that proper analysis and manipulation of big data sets is actually quite difficult. It is easy to find correlation of variables with limited causation.

The proper use of big data analysis will allow mills to use their enormous store of archival data in a positive manner to troubleshoot and solve persistent, long term problems that do not lend themselves to traditional troubleshooting methodologies. Our next step is to apply big data techniques to mill process conditions to predict environmental performance parameters for the mill.



Fu



Hart

Fu is senior technical engineer for MWV Corp. in Covington, VA, USA. Hart is director — Fiber Science for MWV Corp. in Atlanta, GA, USA. Email Hart at peter.hart@mwv.com.