Pulp Machine Sheet Break Reduction

Jon Lindenauer

2016 Fall Technical Conference

Acknowledgement

- Dick Kleinknecht, PhD.
  - Invaluable help with PLS-DA modeling and the model residual control chart.

What Is a Pulp Machine?

A pulp machine uses wood pulp slurry to make a pulp sheet in a continuous process.

We want to run the machine as fast as possible with no sheet breaks.

A Sheet Break Is Bad!

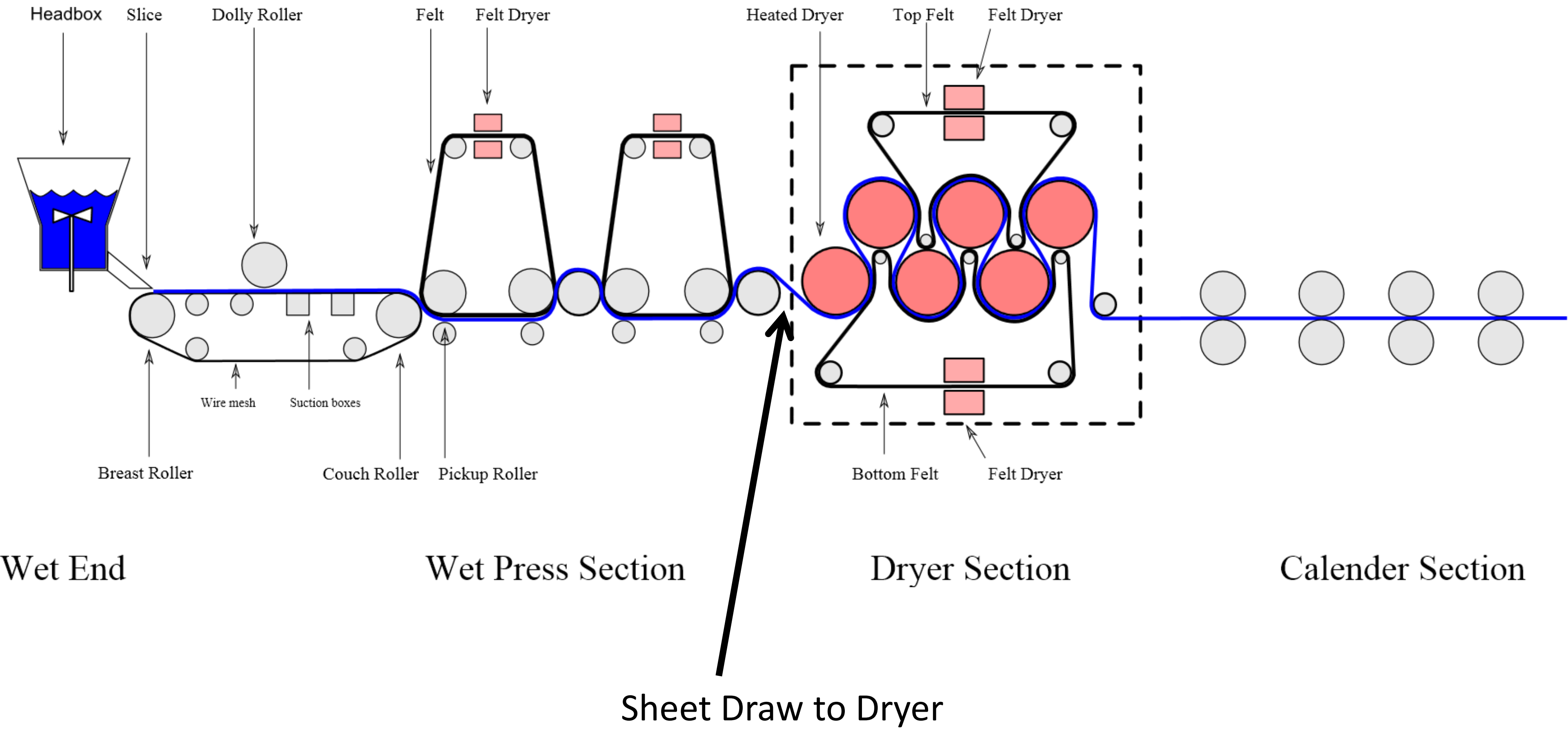
A sheet break stops the continuous production process for 2 to 8 hours.

Lost production costs a lot of money.

Substantially reducing pulp machine sheet breaks could result in millions of dollars of increased revenue and cost avoidance.

[Sheet break video]

Where Breaks Commonly Occur



Why Break at the Sheet Draw?

Pulp machine uses a Flakt dryer with a draw of 30+ feet. A weaker sheet will break in that draw area.

Problem Statement

A softened pulp grade has 7 times more sheet breaks per ton than regular fluff pulp grades.

How can a mill reduce its softened pulp sheet breaks over the course of a year?

If these sheet breaks can be reduced 50%, revenue would increase by many millions of dollars.

Teamwork

A team was formed to solve the pulp machine sheet break problem for the softened grade.

The planned path forward:

1. Gather data to identify/verify the potential causes.
2. Take steps to eliminate the causes.
3. Assess whether causes were eliminated.

Question to the Statistician

How can we use the plethora of mill run-time process data, and the knowledge of when sheet breaks occur, to reduce the break frequency?

Answer: Mining the Plethora of Data

Me, way back when.

Here's my first experience data mining

Getting the Data

How do you reduce hundreds of identified potential variables to a manageable amount?

Meet with the operators and engineers, several times, and ask what do they “look” at when the process is running.

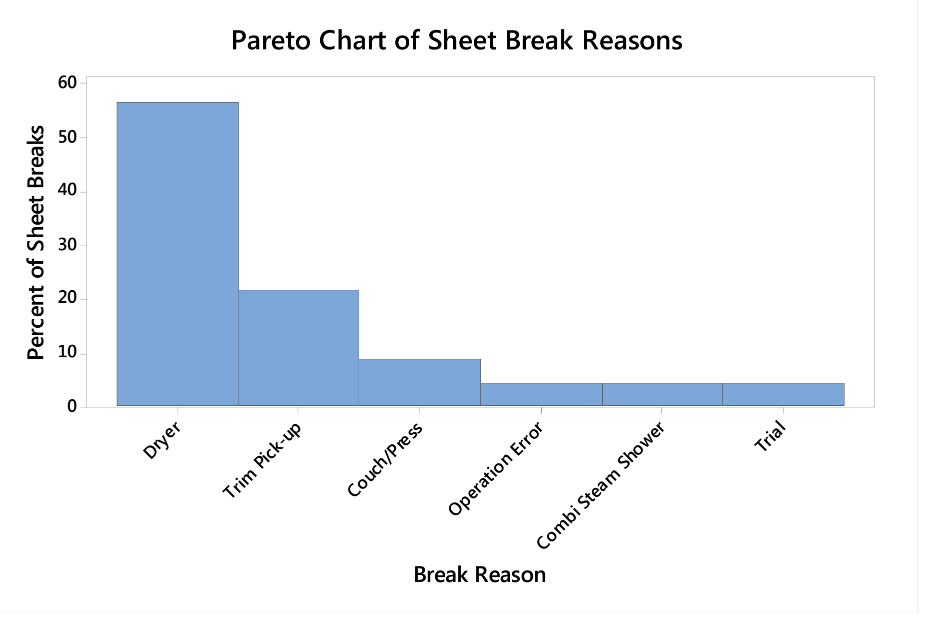
Sit in the control room and watch for a few days.

Research past work on sheet breaks (Mietzke, et.al., 2001).

Find out where the breaks are happening on the machine.

Sheet Break Reason Data

Data collected for the past year showed that most sheet breaks occurred either at the sheet draw or close by at the first roll in the dryer.



The data reduced hundreds of variables to a little less than 70.

Gather Data

The DCS (distributed control system) data is archived in an IP.21 database and accessed with Aspen.

The data can be downloaded with an Excel plug-in, but this tool interpolates or averages the data.

Difficult to get the raw data from this widget.

Used an AspenSQL query to download the actual raw data.

Data Archiving Issue?

The archived mill data may "over compressed".

The dataset should be pre-processed using the "compression factor" (Imtiaz et. al., 2007).

The compression factor (*CF*) is defined as:

$$CF = \frac{number\ of\ original\ measurements}{number\ of\ recorded\ measurements}$$

Recommends that compression turned off for *CF* > 3.

Several tags were over-compressed, this delayed the project for 3 months.

Another Archiving Issue?

Imtiaz also discussed the “quantization factor”.

The quantization factor (*QF*) is an index of a measurement instruments resolution.

The authors recommend that a tag with a *QF* > 0.4 be removed from the dataset, although this is an arbitrary cutoff.

The *QF* is defined as:

$$QF = \frac{minimum\ difference\ between\ consecutive\ points}{standard\ deviation\ of\ the\ tag}$$

All the process data had a *QF* < 0.4.

The *CF* and *QF* were both calculated with Aspen SQL queries.

More Data

3 more months of data were collected in which 4 sheet breaks occurred (**BAD**).

- 5 time spans during this period where chosen where the process was “stable” and there were no sheet breaks (**GOOD**).
- The 3 hours of process data archived at ~1 minute intervals prior to the sheet break, and during a time of a stable process, were used for each time period.

Look for Big Correlation

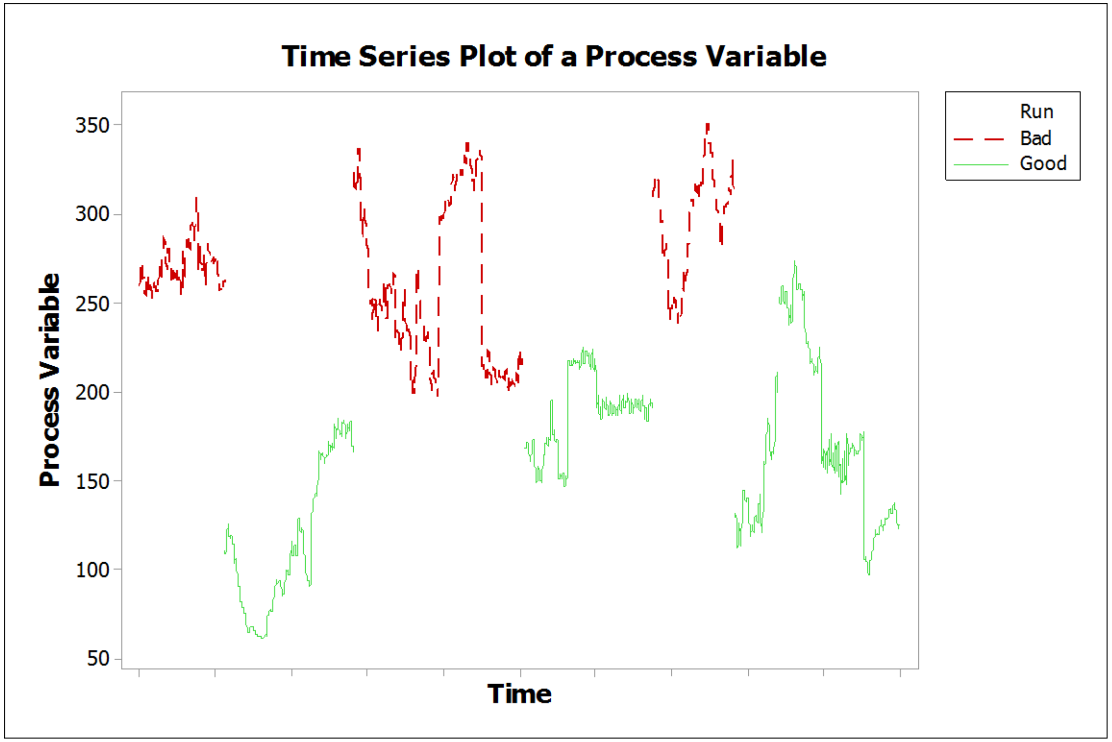
Rule of thumb for big was ***r*** > ~0.85.

Removed tag that operators and engineers deemed more practical to machine operation.

0.772872	0.453481	0.405912	0.522437	-0.130453	-0.624460	0.998937
-0.146286	-0.460189	-0.132770	0.075461	-0.114390	0.306410	-0.134799
-0.137743	-0.492618	-0.145517	0.163118	-0.105753	0.238633	-0.132303
0.023369	0.154255	0.067867	0.433233	0.081463	-0.515186	0.123718
0.143504	0.290925	0.155655	0.196343	-0.329834	-0.579875	0.442188
0.118160	-0.416568	-0.224515	-0.022325	-0.315895	-0.133285	0.386977
-0.035939	-0.215732	-0.026384	0.247307	0.225691	0.147911	-0.246466
1.000000	0.608548	0.572172	0.666139	0.139432	-0.473761	0.775281
	1.000000	0.716299	0.338426	0.157915	-0.515250	0.454078
		1.000000	0.448854	0.056526	-0.437958	0.408872
			1.000000	0.131687	-0.435974	0.525237
				1.000000	0.145391	-0.126731
					1.000000	-0.631919
						1.000000

Clean Up the Data

- More iterations of looking at time series plots of the process data with operators and engineers (including log books). Looked at a lot of plots!



Statistical Methods

We used Partial Least Squares – Discriminant Analysis (**PLS-DA**).

PLS-DA is a multivariate regression analysis that uses group info (e.g., **good** or **bad**) to maximize the separation between **groups** of observations (Eriksson et.al., 2001) using a PCA-like model for the X's and Y's.

PLS-DA will hopefully identify mill variables that are highly predictive of **group** differences.

Statistical Methods

The "**good**" and "**bad**" runs were identified in the process dataset.

Good = no sheet break

Bad = sheet break

Good and **bad** indicator variables are the **Y**'s in the PLS-DA.

Process data variables are the **X**'s.

Statistical Model

PLS-DA model:

[**BAD** **GOOD**] = [lots of process variables]

Can the indicator vectors discriminate (group or separate) the **good** and **bad** runs?

Used ***Simca*** statistical software for multivariate analysis.

Simca centers and scales the X's.

Principle Component Selection

- ***Simca*** will automatically select the number of principle components (PCs) based on maximizing Q^2, the cross-validated percent of variation explained by the model (aka predicted R^2) (Eriksson, et al, 2001).

$$Q^2 = 1 - {}^{PRESS}/_{SSX_{TotCorr}}$$

Variable Selection

Given the number of PCs automatically selected, use the VIP plot to select variables.

VIP - Variable Importance of the Projection (Eriksson, et al, 2001)

The VIP summarizes the importance of the X-variables.

The average VIP is equal to 1, so VIP-values larger than 1 indicate important X-variables.

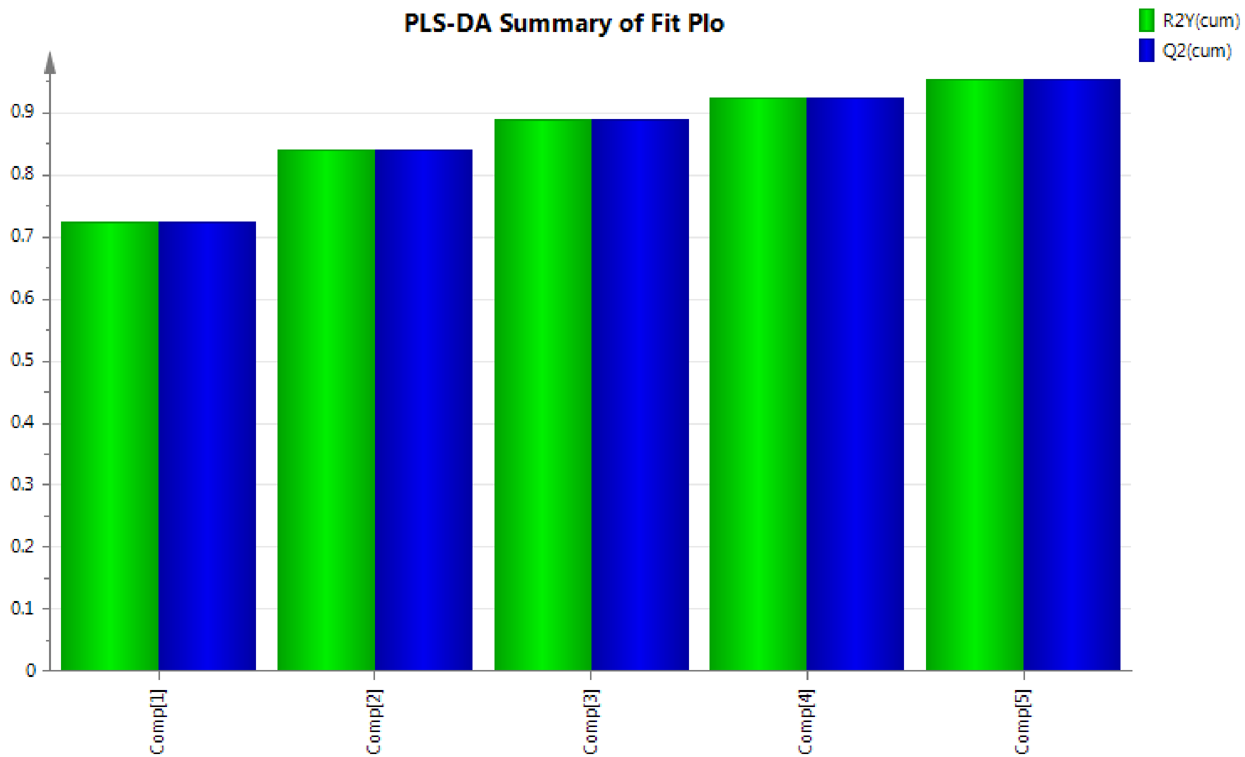
Manual Model Selection

1. Use number of PCs that maximize Q_1^2.
2. Create the VIP plot.
3. Remove any variables with VIP < .5 to .8.
4. Re-calculate number of PCs that maximize Q_2^2.
5. If Q_2^2 is a smaller than Q_1^2, stop. Use previous variables selected and number of PCs.

A simpler model is better if Q_1^2 and Q_2^2 are close.

5 Principle Components Selected

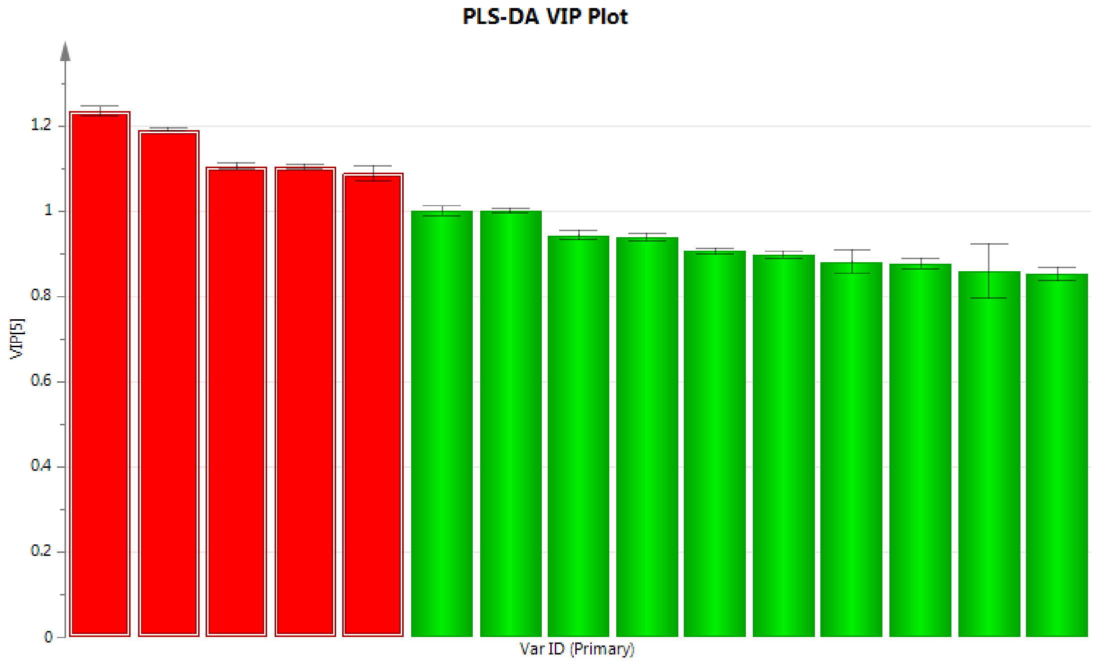
$Q_2^2 = 0.918$ and $Q_1^2 = 0.954$, so stop and use previous variables selected.



15 Variables Selected

15 variables were identified by the PLS-DA VIP plot with VIP > 0.8

Only 5 had VIP > 1.1



DModX Plot

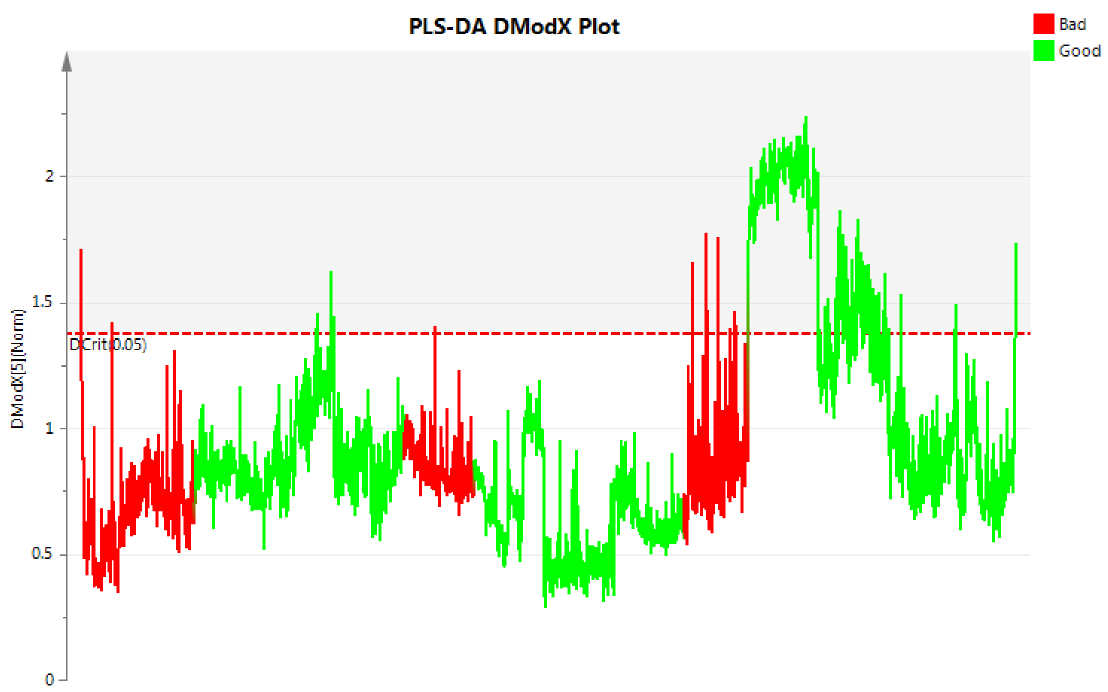
DModX (**D**istance to the **Mod**el of **X**-space) displays how well the observation fits the model.

DModX is displayed in standardized units.

A DModX twice as large as Dcrit identifies moderate outliers. (Eriksson, et al, 2001)

DModX Results

The DModX shows only a few spikes and some of the **good** runs being different from others. No points were more than twice the distance from Dcrit, so we did not remove any of the data points.



Score Plot

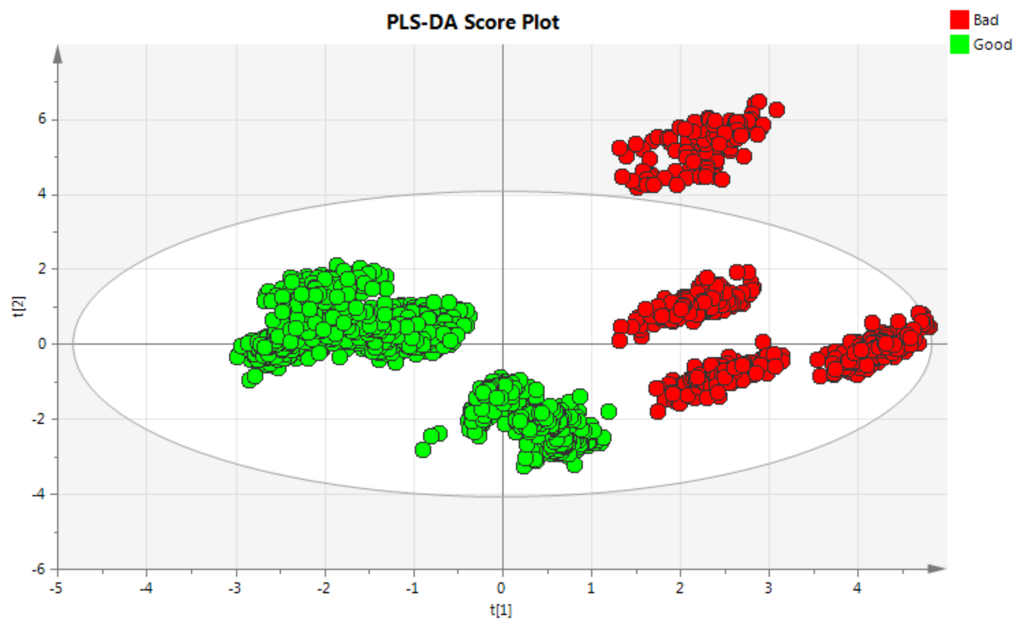
The scores are the new variables summarizing the X-variables. The first score (first principle component) explains the largest correlation of the X's to the Y's, followed by 2nd score, etc.

The score plot shows how the X's are situated with respect to each other, possible groupings and other patterns in the data. (Eriksson, et al, 2001)

PLS-DA Score Plot Results

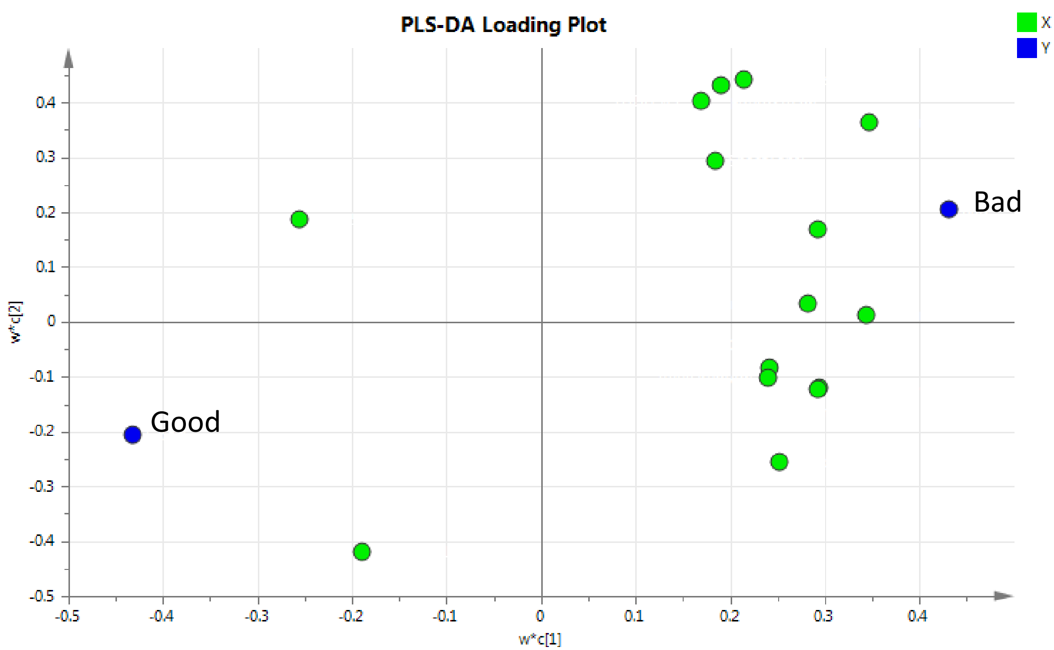
The Score Plot showed that the process variables grouped into "**good**" runs and grouped into "bad" runs.

In other words, the process runs **differently** when a sheet break occurs.

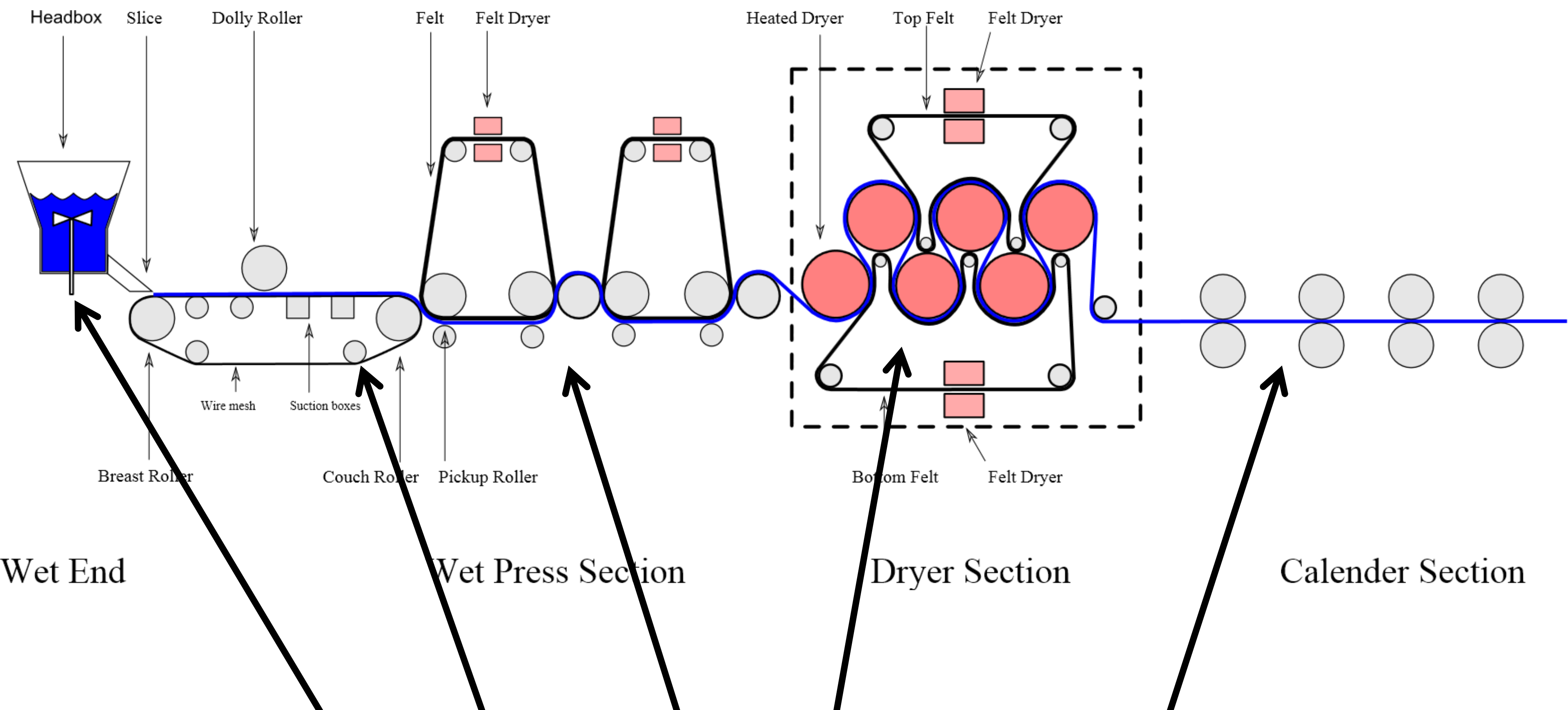


Loading Plot

- The PLS load plot displays the relation between the X's and the Y's. Variables opposite each other are negatively correlated. Variables close to each other are positively correlated. (Eriksson, et al, 2001)



Machine Runs Differently When a Sheet Break Occurs

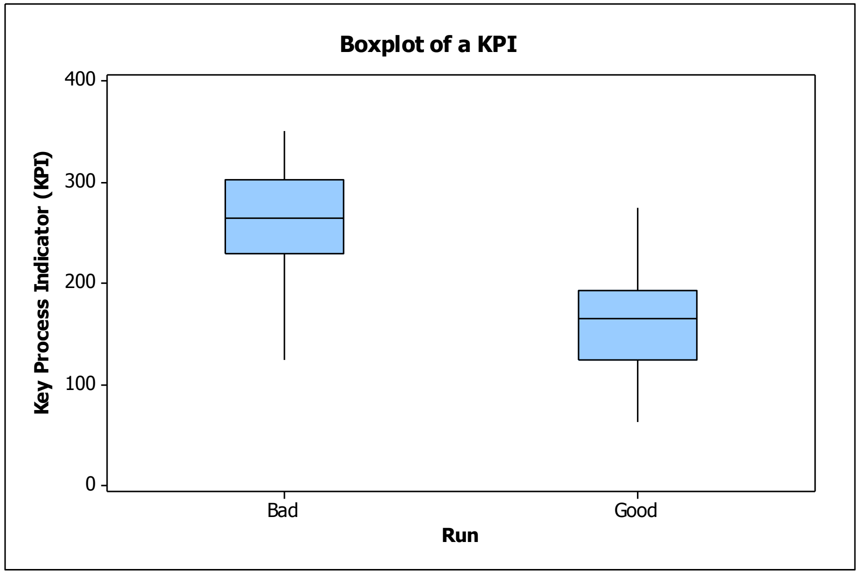


All process areas are part of the model

Five KPIs Identified

The 5 variables with VIP > 1.1 were identified as Key Performance Indicators (KPIs) to be used by the machine operators in a display on their control panel screen.

Box plots showed most of the KPI's process data looked like the example below. Showing that the process variables are different when the machine goes into a sheet break or is running well.



Eliminate Causes

A prediction model using the PLS-DA results of the **GOOD** process data was created with all 15 variables.

Created a control chart of the model residuals that's updated every minute (Jackson, et. al., 1970).

Drill Down

If the residuals control chart exhibits out of control conditions then the operator is instructed to view the KPIs to see which one may be causing the model to deviate from common cause variation.

The Mullen test is a lab test of the sheet strength. This is only done at reel turn up (every ~35 minutes). The Mullen test result is also part of the KPI view.

Time series plot of the KPIs with center lines and 1 and 99 percentile limits have been created.

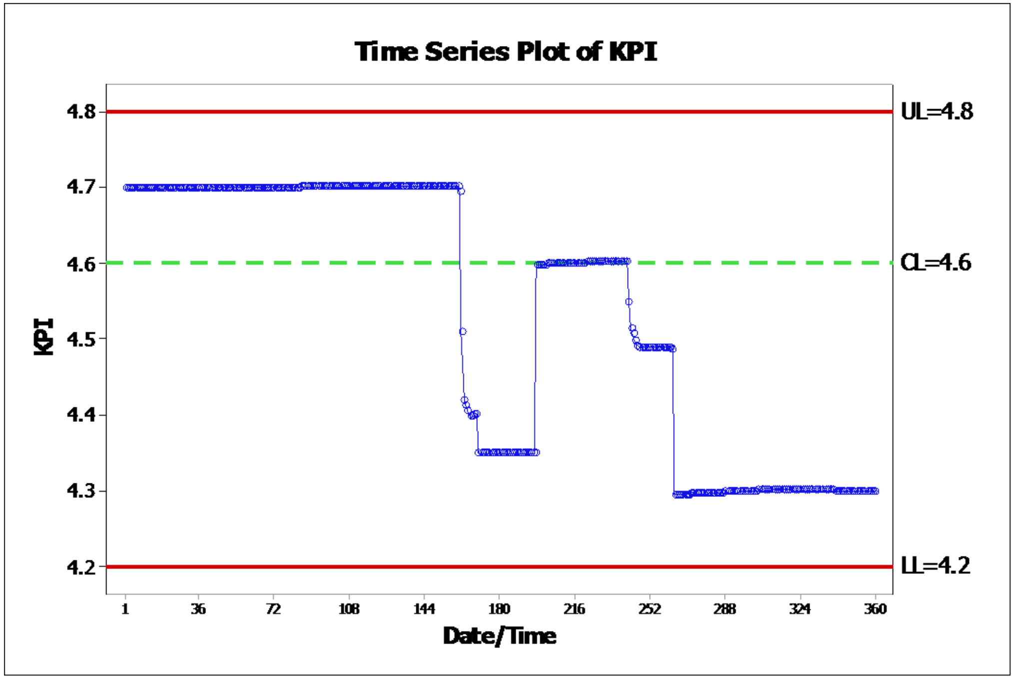
Monitoring the KPIs

Lower limit is 1st percentile

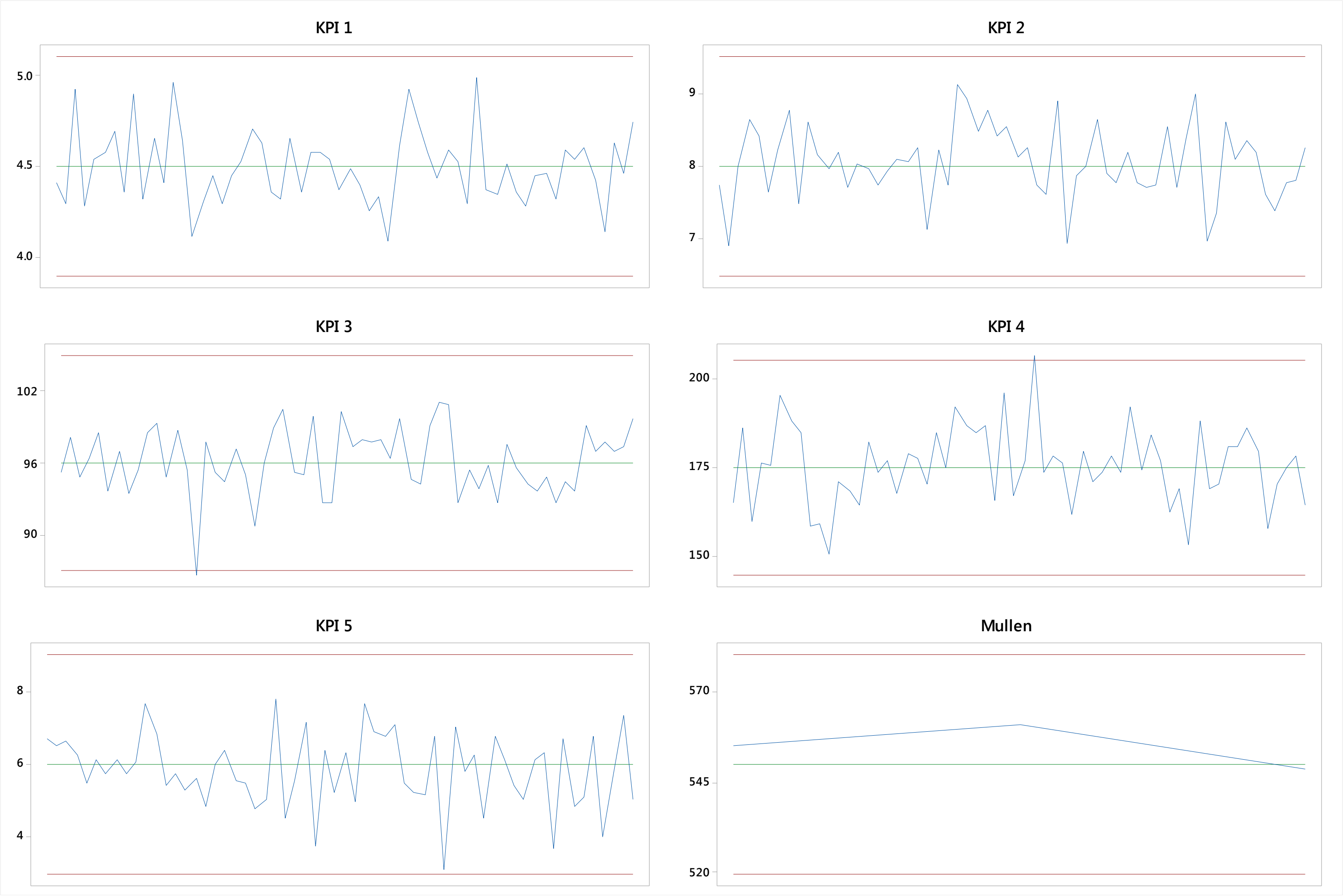
Centerline is 50th percentile

Upper limit is 99th percentile

There is a view of the 5 KPIs and the Mullen test results immediately available to the operators.



Operator Drill Down View



Causes Eliminated?

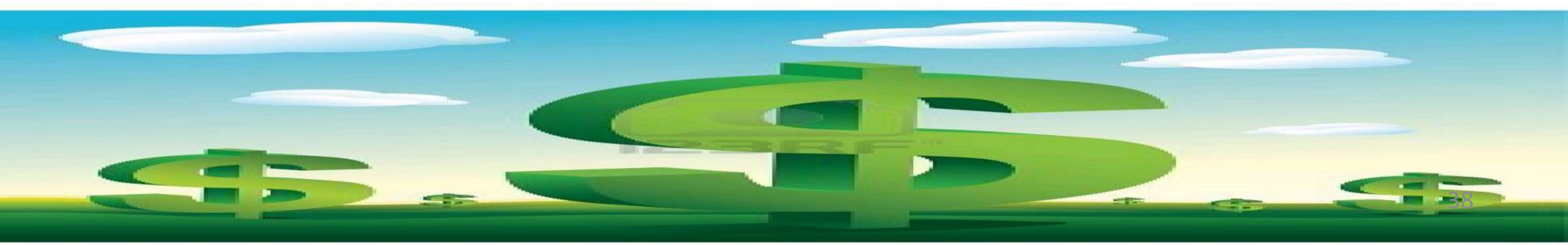
Set a record of 62 days without a sheet break.

A ~65% reduction in softened pulp sheet breaks.

Average machine speed increased from 550 fpm to over 600 fpm for the softened pulp grade.

The increased productivity equates to increased revenue and decreased cost.

The project also supported the sustainable and efficient use of wood and energy resources, as well as a safer work place.



References

Eriksson, L; Johansson, E; Kettaneh-Wold, N; and Wold, S. Multi- and Megavariate Data Analysis: Principles and Applications, Umetrics Academy (2001).

Imtiaz, SA; Shah, SL; Patwardhan, R; Palizban, HA; and Ruppenstein, J. "Detection, Diagnosis and Root Cause analysis of Sheet-Break in a Pulp and Paper Mill with Economic Impact Analysis", The Canadian Journal of Chemical Engineering, 85, pp. 512-525 (2007).

Jackson, JE; and Mudholkar, GS. "Control Procedures for Residuals Associated with Principal Components Analysis", Technometrics, vol. 21 (1979).

Mietzke, DWR; Taylor TT; Shandrow, ER. "Cosmopolis Pulp Machine Diagnostic Survey", Weyerhaeuser Technical Report (2001).

Questions?

Contact Info:

Jon Lindenauer

253.924.4651

jon.Lindenauer@ipaper.com