Stability Analyzer

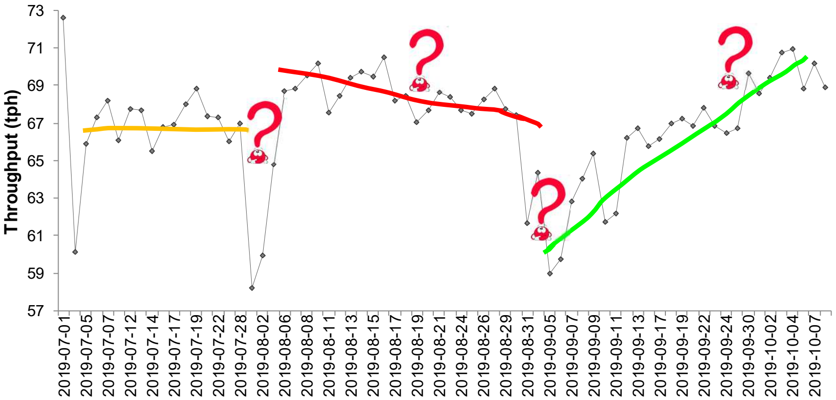
Our add-in just got even better!

The 2023-01-19 Webinar

Introduction

≡ Have you ever wondered :

- Why an important performance parameter (OEE, machine speed, quality) shifted over time?
- Which factors could have contributed to the shifts?

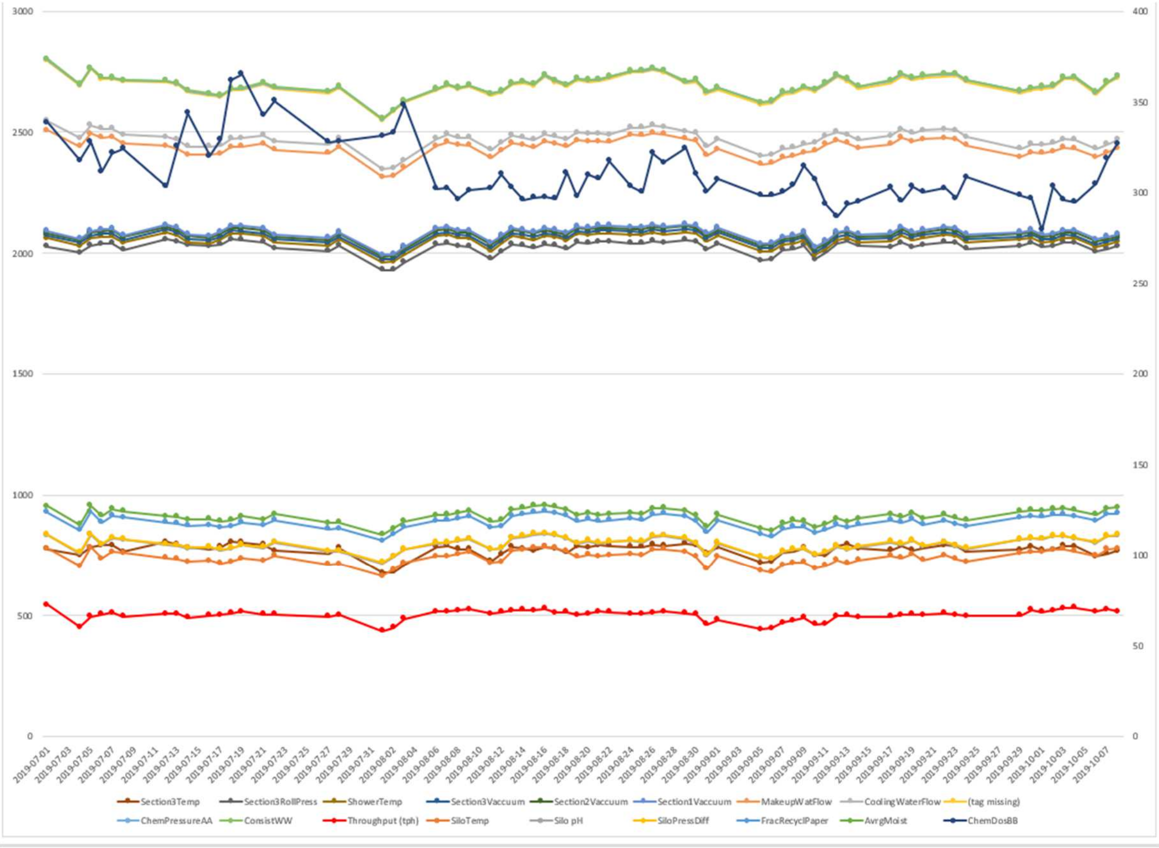


Introduction - Approaches

- Typically, three approaches are used to find the factors that could explain shifts in a performance parameter:
  1) Visually comparing trend charts from each potential factor to the trend of the performance parameter
  2) Performing many simple linear regressions using each potential factor separately
  3) Performing a multivariate analysis (MLR, PLS or other)

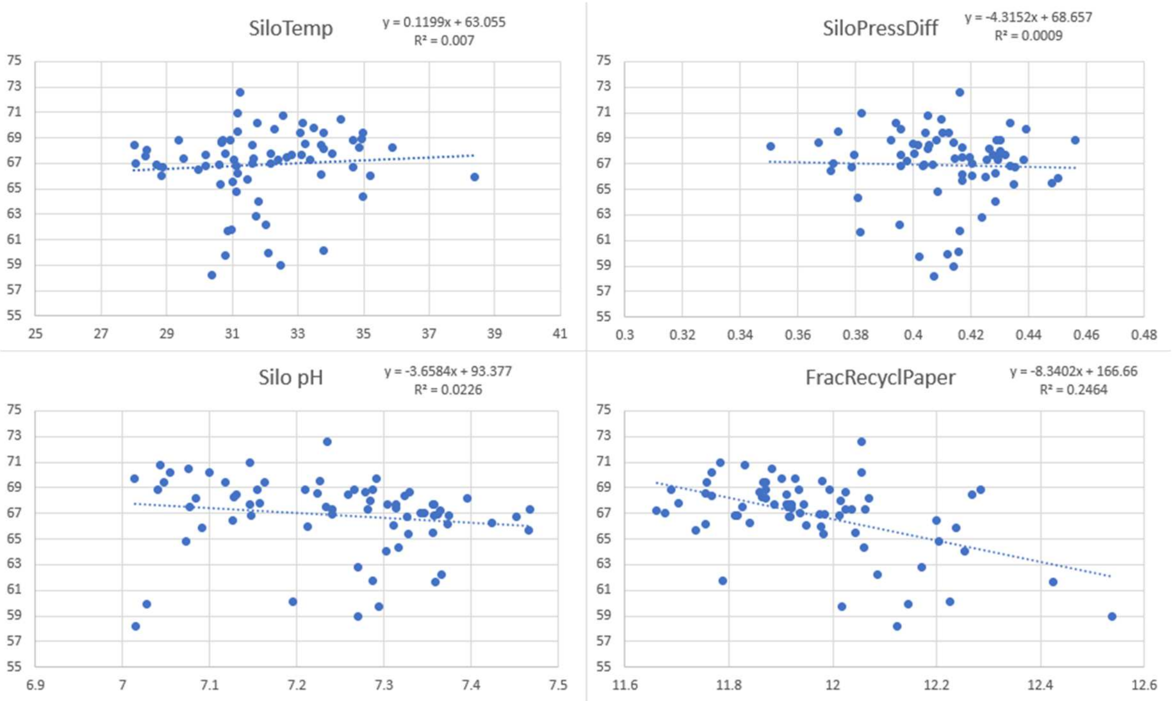
Introduction - Approaches

- Multiple trend charts



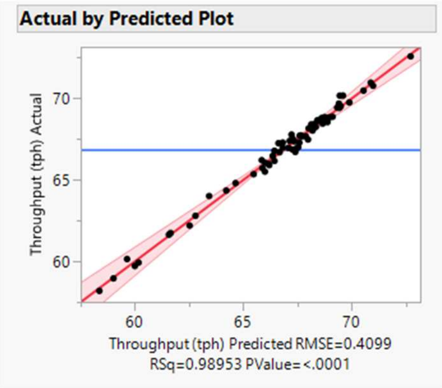
Introduction - Approaches

- Multiple pairwise correlations



Introduction - Approaches

- Multiple linear regression



Effect Summary

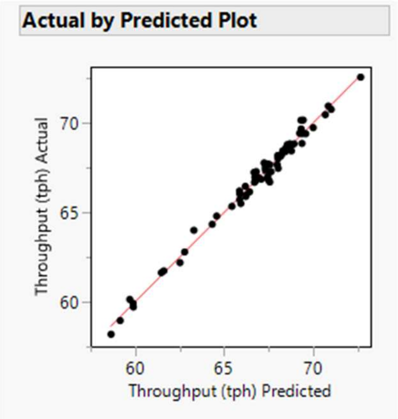
Source	Logworth	PValue
ChemFlowBB	3.646	0.00023
FracRecyclPaper	2.786	0.00164
CoolingWaterFlow	2.772	0.00169
FabricLife	1.728	0.01873
Silo pH	1.706	0.01969
SiloTemp	1.116	0.07659
MainPMsteam	1.029	0.09350
AvrgMoist	0.942	0.11438
ShowerTemp	0.928	0.11805
ConsHB2	0.921	0.11995
SteamPressure	0.892	0.12835
SiloPressDiff	0.874	0.13375
Section1Vaccuum	0.847	0.14223
Section3Vaccuum	0.688	0.20527
WWflow	0.587	0.25883
ConsHB	0.541	0.28806
(tag missing)	0.532	0.29408
ReelSpeed	0.513	0.30691
ConsistWW	0.468	0.34066
MakeupWatFlow	0.426	0.37494
ConsistToPM	0.424	0.37646
Section3RollPress	0.358	0.43838
(tag missing) 2	0.302	0.49920
ChemFlowAA	0.277	0.52887
ChemPressureAA	0.247	0.56673
ChemDosBB	0.189	0.64772
FabricVacRollSpeed	0.092	0.80980
Section3Temp	0.083	0.82689
ReelMotorAmp	0.081	0.82913
Section2Vaccuum	0.063	0.86472
BiocideFlow	0.041	0.90923
Section1Temp	0.039	0.91446

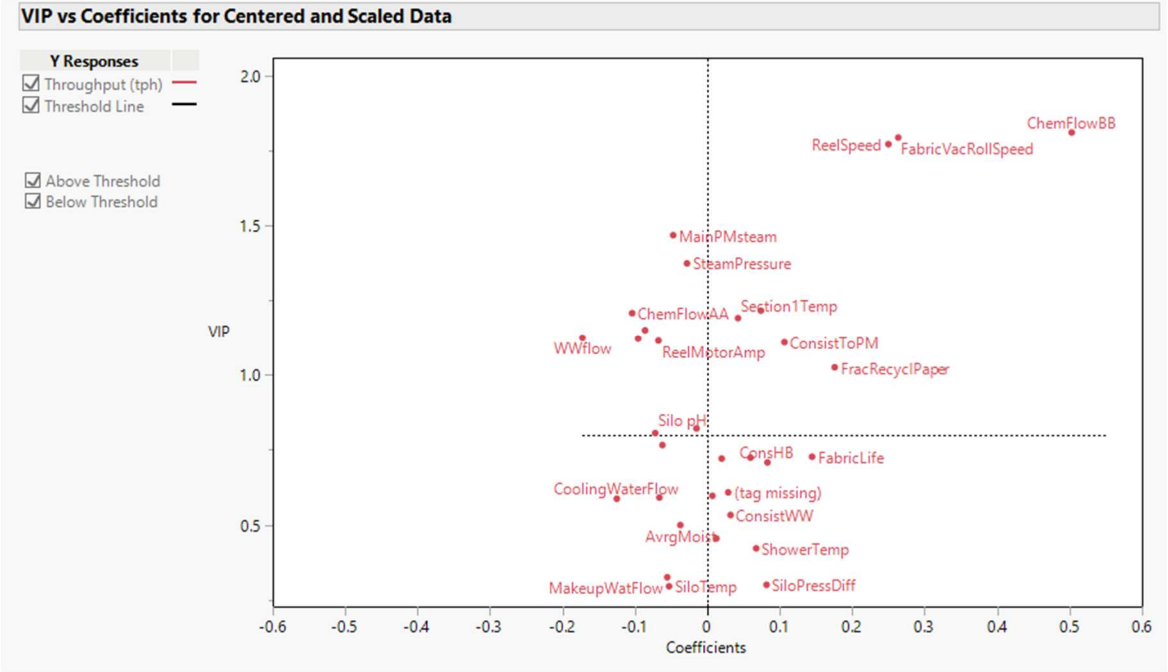
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	103.56275	67.30097	1.54	0.1324	.
SiloTemp	-0.081314	0.044636	-1.82	0.0766	3.4198769
Silo pH	-1.927781	0.790702	-2.44	0.0197*	3.7330205
SiloPressDiff	7.5948666	4.953802	1.53	0.1337	4.4030138
FracRecyclPaper	2.2216495	0.653768	3.40	0.0016*	5.350515
AvrgMoist	-1.455437	0.900102	-1.62	0.1144	4.168425
ChemDosBB	0.0025904	0.005623	0.46	0.6477	4.3877314
Section3Temp	0.0257234	0.116791	0.22	0.8269	3670
Section3RollPress	-0.025135	0.032085	-0.78	0.4384	8.5150452
ShowerTemp	0.0309639	0.019349	1.60	0.1180	8.2724558
Section3Vaccuum	-0.262753	0.203781	-1.29	0.2053	3.8516164
Section2Vaccuum	0.2229969	1.299813	0.17	0.8647	12.167805
Section1Vaccuum	-0.467425	0.311721	-1.50	0.1422	10.843853
MakeupWatFlow	-0.004876	0.005429	-0.90	0.3749	6.1471605
CoolingWaterFlow	-0.18165	0.053642	-3.39	0.0017*	6.3083971
(tag missing)	0.0090253	0.00848	1.06	0.2941	5.0946953
ChemPressureAA	0.0816633	0.141273	0.58	0.5667	13.346338
ConsistWW	0.6140199	0.636086	0.97	0.3407	2.4823331
ConsHB	1.9266462	1.787411	1.08	0.2881	6.6735745
ConsHB2	-3.401651	2.137046	-1.59	0.1199	8.9076546
FabricVacRollSpeed	0.0008996	0.003711	0.24	0.8098	280.25166
ConsistToPM	1.7052172	1.9048	0.90	0.3765	13.725524
BiocideFlow	0.0029645	0.025824	0.11	0.9092	4.8378539
Section1Temp	-0.012625	0.116732	-0.11	0.9145	3721.7474
MainPMsteam	-0.002963	0.001721	-1.72	0.0935	935.43712
SteamPressure	-0.218986	0.140786	-1.56	0.1284	649.41124
ReelSpeed	-0.004635	0.004474	-1.04	0.3069	443.06267
FabricLife	0.0439693	0.01788	2.46	0.0187*	9.4723733
(tag missing) 2	-0.009673	0.014174	-0.68	0.4992	70.591469
ReelMotorAmp	0.0002827	0.001301	0.22	0.8291	22.30368
WWflow	-0.011843	0.010327	-1.15	0.2588	92.382156
ChemFlowBB	0.0166493	0.004074	4.09	0.0002*	269.51886
ChemFlowAA	-0.023369	0.03676	-0.64	0.5289	28.404189

Introduction - Approaches

- PLS





Introduction - Approaches

≡ Some drawbacks of usual approaches:

1) **Visual trends:** Tedious, arbitrary, no quantification of strength of relationship, no visibility on interactions
2) **Simple linear regressions:** Tedious with Excel but fast with JMP, a lot of weight is put on extreme conditions, not robust to outliers, no visibility on interactions, link between the shifts in performance and the correlation not always obvious (did the change happened at the same time on both parameters?)
3) **Multivariate analysis:** a lot of weight is put on extreme conditions, not robust to outliers, link between the shifts in performance and the correlation not always obvious (did the change happen at the same time on both parameters?)

Introduction

How it works

Why?

- Because:
  - Lack of structured and methodic procedure to identify stable process periods and possible factor contributing to the changes
  - Historical data comes with noise, outliers, changes in means,…
  - Finding the different stable process periods is a challenge
  - Finding the underlying signals matching the different stable periods is another challenge
  - Need to be able to analyze with minimum data preparation
- Also because:
  - After 20 years spending hours trying to accomplish this task, we are getting lazy and starting to know how we answer this!

Overview

1) Change Point Analysis on Y (find the CPs)
   - Determine the stable periods of the process
2) Changes in medians around the CPs for Y and all Xs
   - Robust local "how much variation" (delta) every time Y changes
3) Simple regressions on the deltas between each X and Y
   - Get the correlation between the stable periods medians
   - One pair of ΔX and ΔY for each CP, get the slope (=average change)
   - Count matching patterns in the medians profile changes
4) Post-processing of the information
   - Filter out insignificant Xs (regression p-value too high)
   - Calculate stable period medians and plot them on trend charts
   - Identify potentially correlated factors to significant Xs

1) Identify stable periods

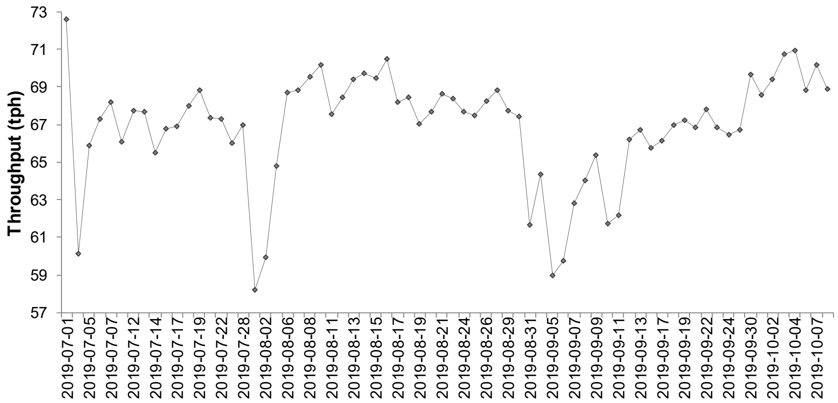
- Let's consider the Y variable as the "main" performance measurement
- There exists a powerful and proven tool : the Change Point Analysis (CPA)

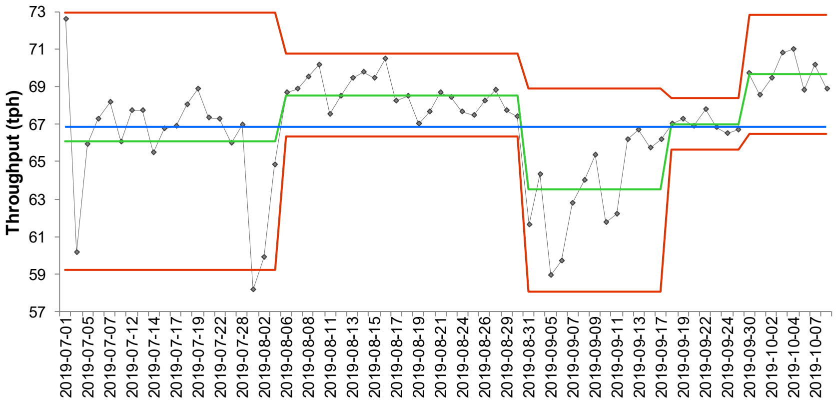
  https://variation.com/wp-content/uploads/change-point-analyzer/change-point-analysis-a-powerful-new-tool-for-detecting-changes.pdf

  - Based on CUSUMs to determine the moment of the changes
  - Also based on bootstrapping to estimate the chances that a change is "true" (too special to be just luck)
  - It is robust to the presence of outliers, to non-Normal distributions, and to mild autocorrelation
- Result: list of Change Points (CP) delimiting the different periods when the process was stable

1) Identify stable periods

From	To	p-value	Average
2019-07-01	2019-08-03		66.11
2019-08-06	2019-08-30	0.0005	68.55
2019-08-31	2019-09-17	0.0000	63.50
2019-09-18	2019-09-29	0.0006	66.98
2019-09-30	2019-10-08	0.0018	69.66

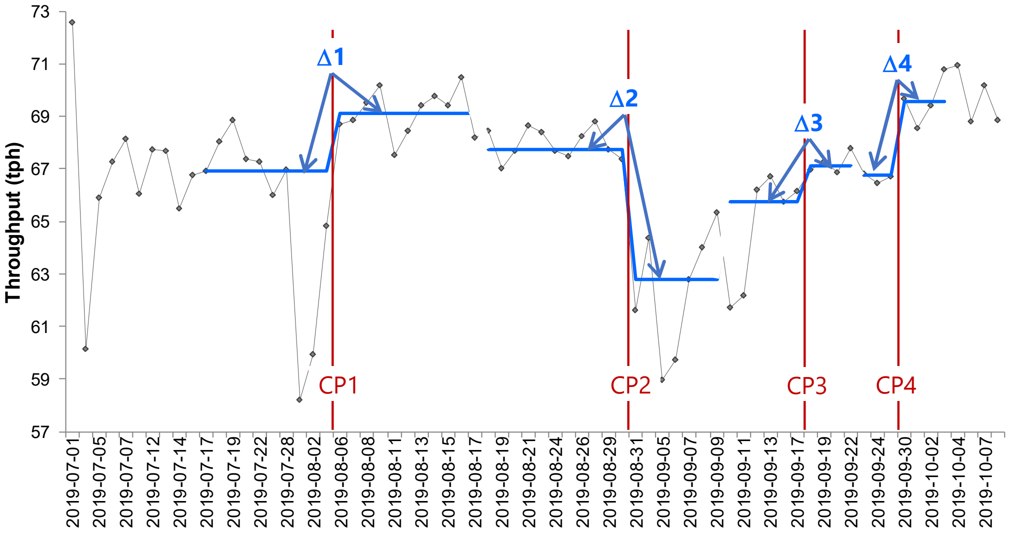




2) Changes around the CPs

- General idea:
  - How much did the process change between the stable periods?
  - How much did each X change at the same moments?
  - We will be using medians, not means:
    - Because they are more robust to outliers!
    - To equalize the CP weights regardless of segment lengths
- The math:
  - For a CP of a variable:
    - Calculate the median of the data from the mid-previous period to the CP (median before) and from the CP to the mid-new period (median after)
    - Compute delta = (median after) – (median before)
  - Apply to all variables and for all CPs

2) Changes around the CPs



Change points	Median Before	Median After	Delta
CP1	66.90	69.13	2.23
CP2	67.73	62.80	-4.93
CP3	65.73	67.12	1.39
CP4	66.77	69.55	2.78

2) Changes around the CPs

- Applying this procedure on Y and on all Xs we obtain a summary table similar to this one :

	Y	X1	X2	X3	X4	X5
Change Point	Delta(Throughput (tph))	Delta(ChemFlowBB)	Delta(FabricVacRollSpeed)	Delta(ReelMotorAmp)	Delta(Silo pH)	Delta(Section3Vaccuum)
2019-07-01	2.23	129.79	168.90	-72.58	-0.1168	0.0559
2019-08-06	-4.93	-362.57	-460.13	50.44	-0.0010	0.2909
2019-08-31	1.39	100.99	77.18	31.20	-0.0186	0.0097
2019-09-18	2.78	162.49	55.47	-103.74	0.0407	-0.0777

- And we are ready for the next step!

3) Regression and pattern matching

≡ General idea:

- It is possible to correlate the medians of the segments
  - This indicates if in general X and Y are correlated
  - It also provides a p-value…

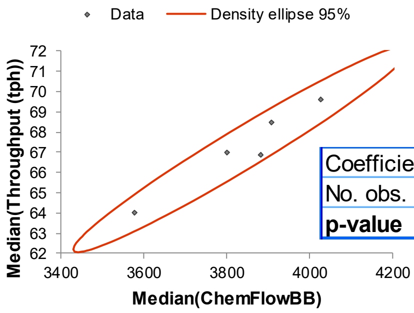
- Given the previous table, when a X changes, how much does the Y change? Simple linear regression!
  - The slope of the regression line gives the average $\Delta Y/\Delta X$
  - This quantity represents: by how much does Y change when X changes of 1 unit?

3) Regression and pattern matching

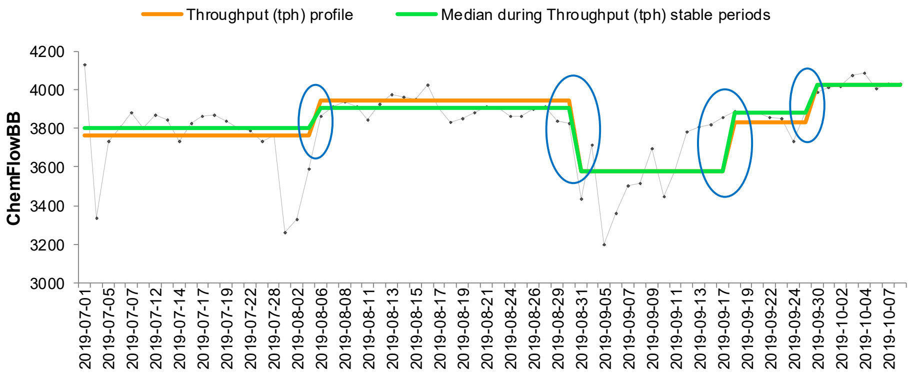
- General idea:
  - Another approach, non-parametric this time: how often does X change in the expected direction when Y changes?
    - Only when ΔX is > $\frac{1}{4}\sigma_{MR}$ (the ImR control chart sigma)
    - Consider the sign of the regression slope
    - If the moves (+ or -) on Y matches the move on X (or the inverse move if the regression slope is negative), add 1 point; return Similarity = 100*points/number of Change Points
      - 100% = perfect pattern match
      - 0% = they never move the same way…
    - Referencing Y change points because they are statistically proven!

3) Regression and pattern matching

Segments	Median(Throughput	Median(ChemFlowBI
2019-07-01 --> 2019-08-03	66.93	3799.52
2019-08-06 --> 2019-08-30	68.46	3906.53
2019-08-31 --> 2019-09-17	64.01	3577.50
2019-09-18 --> 2019-09-29	66.87	3880.24
2019-09-30 --> 2019-10-08	69.55	4024.52

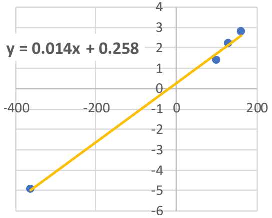


Coefficient r (Pearson)	0.9727
No. obs.	5
p-value	**0.0054**



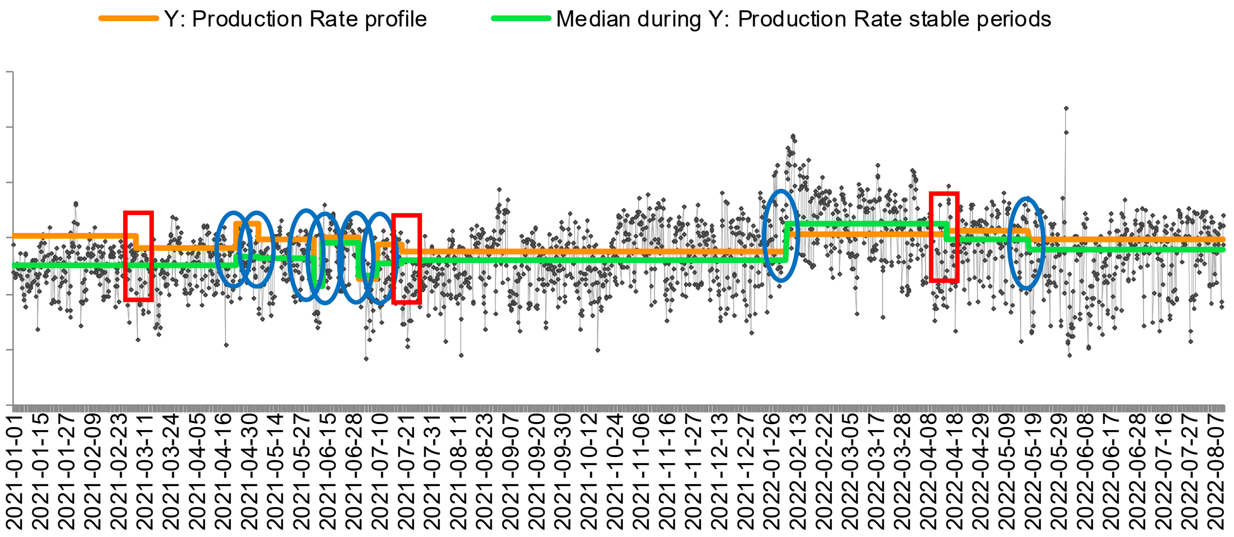
Similarity = 4/4 = 100%

Change Point	Delta(Throughput (tph))	Delta(ChemFlowBB)
2019-07-01	2.23	129.79
2019-08-06	-4.93	-362.57
2019-08-31	1.39	100.99
2019-09-18	2.78	162.49



3) Regression and pattern matching

≡ An example of good but imperfect pattern matching...



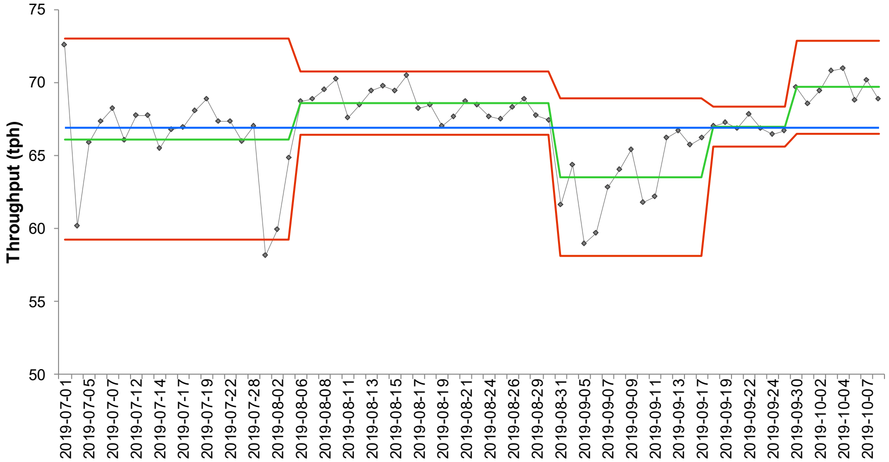
Similarity = 8/11 = 73%

4) Post-processing of the information

≡ Several user-friendly add-ons

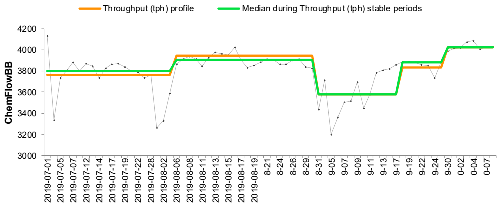
- Results ordered by descending order of Similitudes
  - Ties are broken by ascending p-values
- Filtering out of the Xs that are not statistically significant
  - "Too high" p-values are optionally hidden from the results
- Correlations between Xs
  - Warning that some Xs could be correlated… each X was modelled against Y separately!
- Visual assessment: superposing the Y "profile" over the X medians during the stable periods
  - Do they visually evolve similarly?

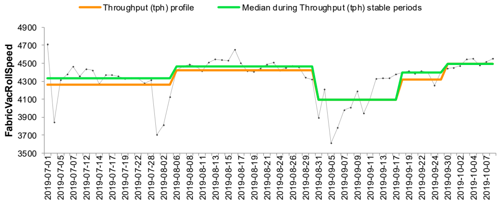
The final result…

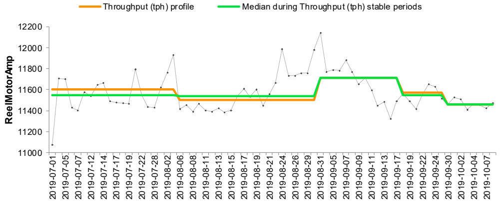


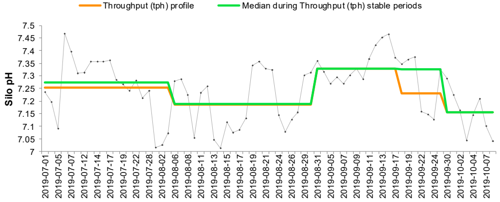
Segments summary			CPA [5000 ¤ 0.75 ¤ 6 ¤ 10000 ¤ 0.95]		
From	To	p-value	Average	No. obs.	Std. dev.
2019-07-01	2019-08-03		66.1066686	20	3.30277474
2019-08-06	2019-08-30	0.0002	68.5491881	22	0.9477753
2019-08-31	2019-09-17	0.0000	63.504098	13	2.54515449
2019-09-18	2019-09-29	0.0004	66.9848793	7	0.4246268
2019-09-30	2019-10-08	0.0020	69.6558802	8	0.91493396

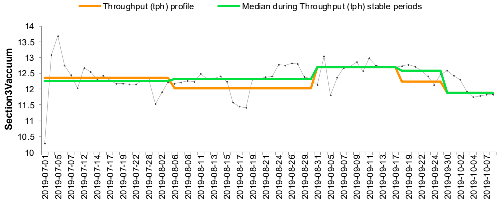
Variable (click on it!)	r (medians)	›-value of	ΔY / ΔX	Similarity	*Correlated variables*
ChemFlowBB	0.973	0.0054	0.0144067	100.0%	*FabricVacRollSpeed; ReelMotorAr*
FabricVacRollSpeed	0.967	0.0072	0.0122134	100.0%	*ReelMotorAmp; ChemFlowBB;*
ReelMotorAmp	-0.959	0.0099	-0.035892	75.0%	*FabricVacRollSpeed; ChemFlowBE*
Silo pH	-0.875	0.0520	-9.952684	50.0%	
Section3Vaccuum	-0.867	0.0572	-21.7096	50.0%	











To conclude...

Final words

- The procedure to identify important Xs added to the Différence's Excel add-in has many advantages:
  - Fast (even if >250 Xs!) and robust to outliers when the main goal is to explain significant changes in a process mean
  - Combines two approaches to ensure even more robust conclusions (regression on deltas and change direction concordance)
  - Great for a rapid screening of many factors
  - Results are easy to interpret and the graphs are easy to understand

- Of course it does not replace advanced statistical analysis (it does not consider interactions or non-linear patterns) but it is still a good first step for screening

différence
consulting in statistics

Différence is a society offering coaching, consulting and training services in statistic, data science, simulation and continuous improvement.

We promote the use of quantitative tools that can be applied at the different steps of an improvement and variability reduction project.

Powerful methods

Adapted approach

Combining hard work with fun

For more information, you can contact:

Martin Carignan, M.Sc., MBA
Associate, Statistician
+1 (514) 795-8000
mcarignan@difference-gcs.com
linkedin.com/in/martincarignan

Vincent Béchard, MASc.
Associate, Analytical Decision Specialist
+1 (438) 521-5829
vbechard@difference-gcs.com
linkedin.com/in/vincentbechard