Large-Scale Process Monitoring using JMP®

Process Monitoring and Diagnosis by Multivariate Statistical Process Control (MSPC)

Jianfeng Ding, Senior Research Statistician Developer
JMP Division, SAS Institute

Table of Content

- **Process Monitoring and Diagnosis**
  - Purpose
  - Data
- **Multivariate Statistical Process Control (MSPC)**
  - PCA / PLS
- **Process Application**
  - Low Density Polyethylene Process (LDPE)
  - Tennessee Eastman Process
- **Demos and Conclusions**

Process Data Analysis - Purposes

- Monitoring the state of the process
  - Early detection
  - Diagnosis and adjustment

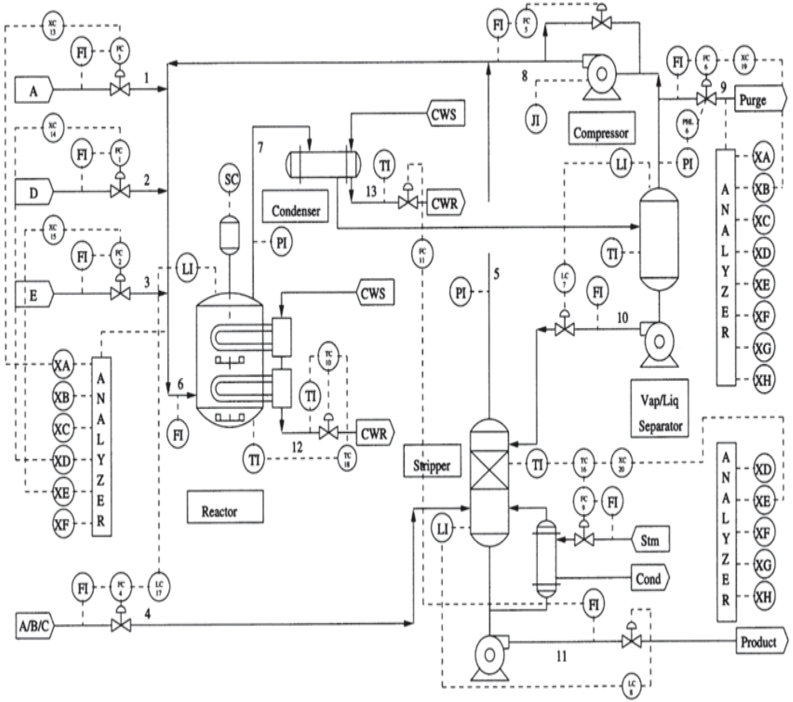
- Understanding the relationship between
  - Input variables, X (process data) and output data, Y (product quality, cost, amount, …)

- Optimization
  - Use process models to improve process

Challenge

- A few decades ago
  - Few variables

- Today
  - Many measurements
  - Large data sets
- Process the same
- Data have changed
  - p = 5 -> 500
  - n = 10 -> 1000



Process diagram

Traditional Statistical Process Control (SPC)

- SPC charts
  - Shewhart, CUSUM and EWMA

- Disadvantage
  - Charts a small number of variables
  - Examines them one at a time

- Most outliers remain undetected
  - No covariance information

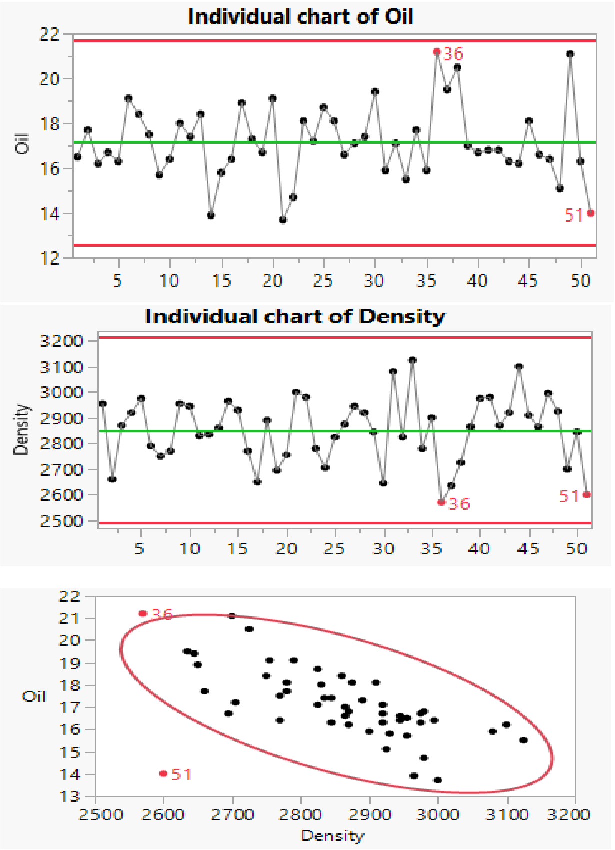


Table of Content

- Process Modelling and Diagnosis
  - Purpose
  - Data
- **Multivariate Statistical Process Control (MSPC)**
  - PCA / PLS
- **Process Application**
  - Low Density Polyethylene Process (LDPE)
  - Tennessee Eastman Process
- **Demos and Conclusions**

PCA and PLS Projection Methods

- **PCA finds the latent variables**

  $$X = TP^T + E \qquad T = XP$$

  - Best explain the variance in X

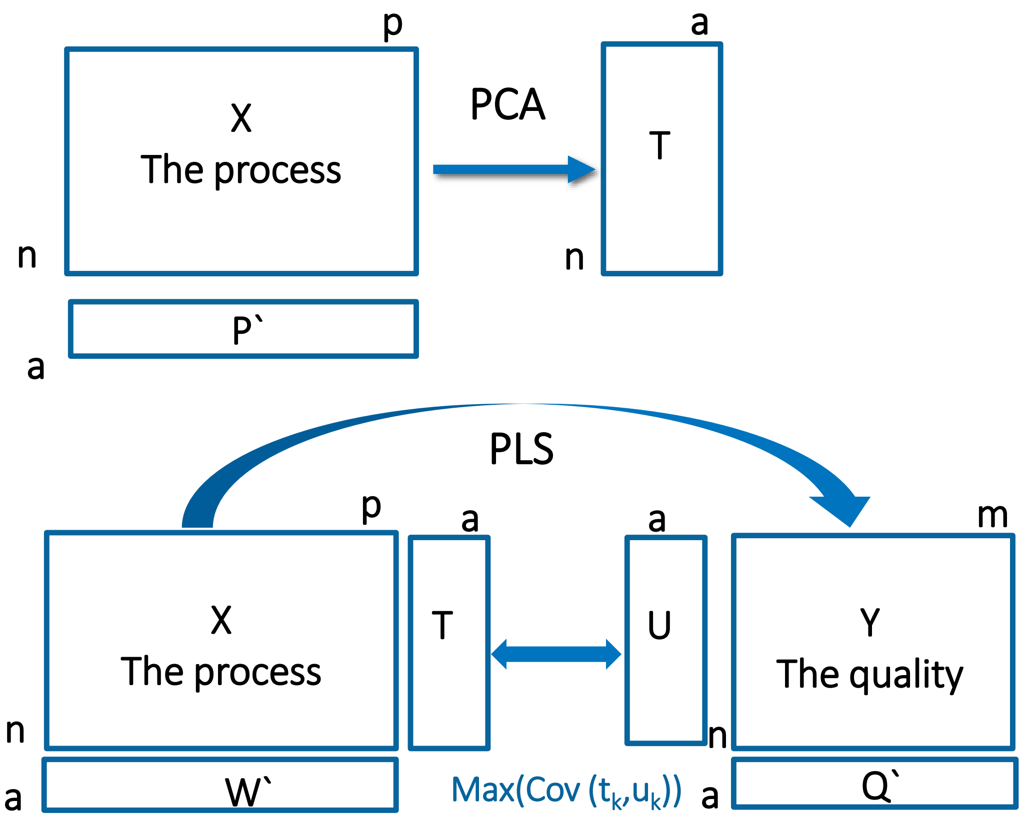
a << p

- **PLS finds the latent variables**

  $$X = TP^T + E \qquad T = XR$$

  $$Y = TQ^T + F$$

  - Best explain the variance in Y
  - Have the greatest relationship with Y



Multivariate Statistical Process Control (MSPC)

Projection methods

Build models

Historical process quality data

PCA and PLS	
Advantages	• Handles a large number of variables • Reduces dimensionality • Handles correlation
Monitoring processes	• Hotelling's T^2 • Squared prediction error (SPE) • Scores
Diagnosing processes	• Contribution plots

Apply models

New process data

Table of Content

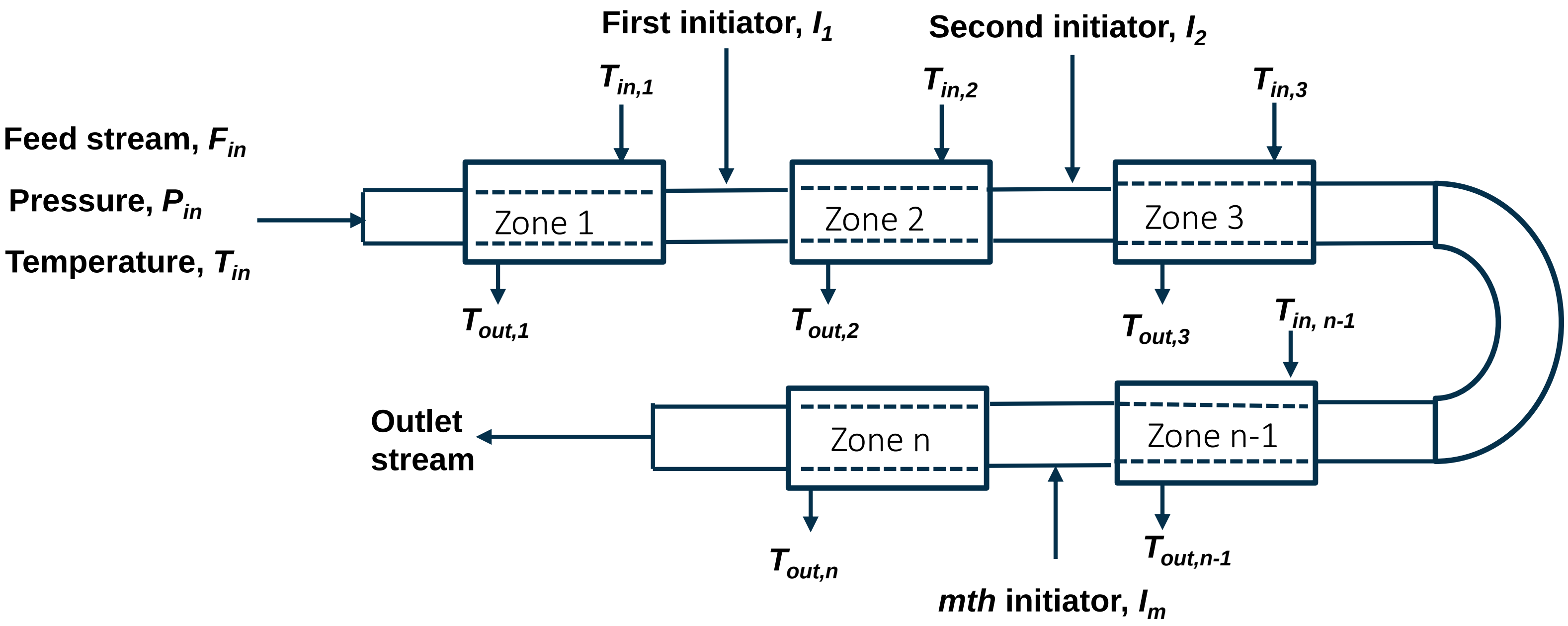
- Process Modelling and Diagnosis
  - Purpose
  - Data
- Multivariate Statistical Process Control (MSPC)
  - PCA / PLS
- **Process Application**
  - Low Density Polyethylene Process (LDPE)
  - Tennessee Eastman Process
- **Demos and Conclusions**

LDPE – Low-density polyethylene

LDPE is a thermoplastic made from petroleum. It was the first grade of polyethylene, produced in 1933 using a high pressure process via free radical polymerization. Its manufacture employs the same methods today.

LDPE is widely used for manufacture various containers, squeezable bottles, wash bottles, tubing, plastic parts for computer components. Its most common use is in plastic bags.

A High-pressure Tubular Reactor



Length of tubular reactor ranges from 500 to 1500m

Data Overview

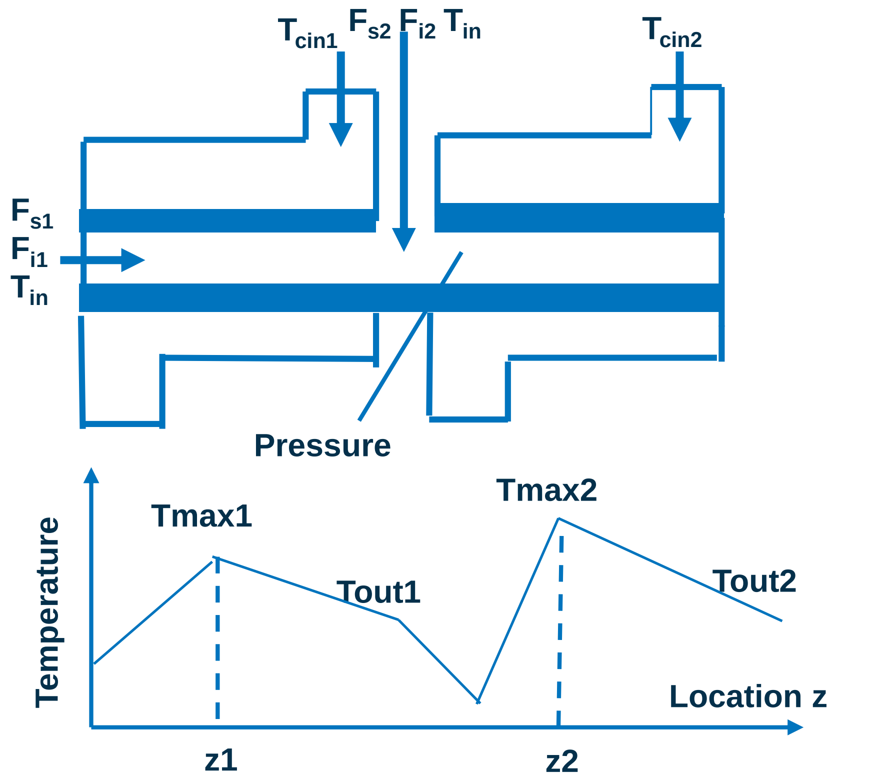
The simulated data based on [MacGregor et al (1994)]

	Tin	Tmax1	Tout1	Tmax2	Tout2	Tcin1	Tcin2	z1	z2	Fi1	Fi2	Fs1	Fs2	Press	Conv	Mn	Mw	LCB	SCB
91	206.3	297.4	231.2	284.6	243.6	116.6	118.0	0.0...	0.5778	0.47...	0.46...	659.0	245.8	2995	0.134	27210	165711	0.8082	26.19
92	206.6	293.8	229.9	280.9	242.5	116.9	118.5	0.0...	0.5862	0.42...	0.40...	669.2	245.9	2991	0.129	27828	162148	0.7615	25.64
93	207.8	297.3	234.5	283.3	241.5	117.1	117.8	0.0...	0.5795	0.48...	0.45...	661.9	245.1	2974	0.133	27195	160547	0.8025	26.30
94	206.8	296.9	237.2	281.1	238.9	117.6	117.5	0.0...	0.5869	0.47...	0.43...	667.1	245.7	2995	0.132	27427	161954	0.7845	25.91
95	205.2	297.6	230.4	284.6	241.8	116.4	117.6	0.0...	0.5794	0.49...	0.46...	666.9	244.9	3016	0.135	27247	166233	0.8094	26.02
96	206.8	[illegible]	229.0	284.1	240.0	116.4	117.3	0.0...	0.5816	0.46...	0.48...	[illegible]	248.6	3004	[illegible]	27267	163262	0.7956	[illegible]
97	205.7	[illegible]	237.5	283.2	[illegible]	[illegible]	[illegible]	[illegible]	0.5810	0.46...	0.43...	[illegible]	246.2	3021	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
98	206.7	[illegible]	230.3	286.7	245.7	116.6	118.1	0.0...	0.5761	0.42...	0.47...	[illegible]	246.8	3043	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
99	209.0	294.9	231.4	284.2	239.4	116.9	117.2	0.0...	0.5795	0.45...	0.47...	676.1	247.1	2997	0.132	27366	161752	0.7876	26.18
100	207.6	296.5	233.8	284.7	243.0	117.1	117.9	0.0...	0.5776	0.46...	0.47...	659.6	245.9	2984	0.133	27244	165684	0.8058	26.20
101	207.7	296.6	233.2	284.8	241.0	117.0	117.4	0.0...	0.5794	0.47...	0.48...	670.3	243.8	3004	•	•	•	•	•
102	206.2	296.1	233.0	286.1	244.6	117.0	117.9	0.0...	0.5763	0.45...	0.48...	670.7	247.0	3020	•	•	•	•	•
103	206.9	295.7	232.7	286.2	244.2	117.0	117.9	0.0...	0.5730	0.46...	0.48...	659.3	244.4	2997	•	•	•	•	•
104	206.7	297.0	232.2	285.4	241.0	116.8	117.3	0.0...	0.5775	0.46...	0.48...	663.0	246.5	3017	•	•	•	•	•
105	207.1	297.9	235.3	280.9	242.0	117.1	118.3	0.0...	0.5877	0.48...	0.41...	663.4	248.0	3017	•	•	•	•	•
106	204.6	[illegible]	226.4	279.8	239.6	116.3	118.0	0.0...	0.5877	[illegible]	0.39...	667.8	246.9	2990	•	•	•	•	•
107	206.7	[illegible]	227.7	284.4	[illegible]	[illegible]	117.4	0.0...	0.5772	[illegible]	0.46...	668.4	243.6	2989	•	•	•	•	•
108	205.5	[illegible]	227.8	285.6	243.5	116.6	117.8	0.0...	0.5789	[illegible]	0.48...	665.4	246.6	3013	•	•	•	•	•
109	206.9	295.4	233.9	284.8	245.0	117.3	118.3	0.0...	0.5766	0.46...	0.47...	655.9	241.5	2977	•	•	•	•	•
110	209.0	295.3	232.1	284.6	241.2	117.0	117.5	0.0...	0.5812	0.43...	0.47...	670.5	244.5	3018	•	•	•	•	•
111	206.8	295.1	228.4	287.2	243.8	116.4	117.6	0.0...	0.5728	0.45...	0.50...	672.0	245.1	2996	•	•	•	•	•

Process

Quality

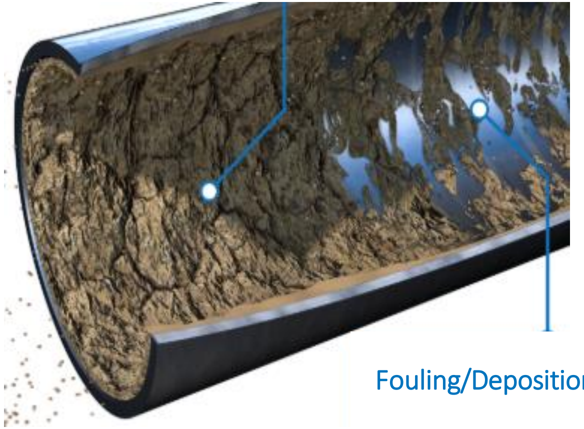
New



Two-zone LDPE reactor with a typical temperature profile

Common Problems

- **Common problems that affect LDPE process**
  - Impurity contamination **(affect temperature profiles)**
  - Change in initiator efficiencies
  - Fouling (deposition) of the reactor walls
  - Equipment, sensors, operators and lab analysis

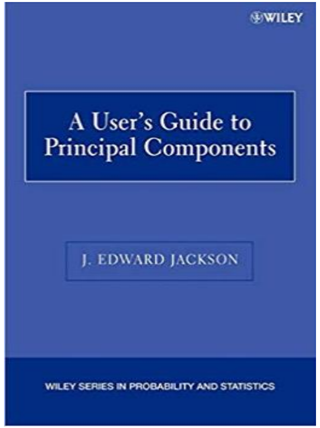


Fouling/Deposition

- **Affect polymer's quality**
- **Affect the behavior of the polymer in its final application**

Multivariate Process Control Procedure

- **Four conditions (quotes from Jackson 1991):**
  - A single answer should be available to answer the question: “Is the process in control?”
  - An overall Type I error should be specified.
  - The procedure should take into accounts the relationships among the variables.
  - Procedures should be available to answer the question: “If the process is out-of-control, what is the problem?”
- **Let’s check model driven multivariate control chart**



Demos

1. PLS modeling historical data
2. T2, SPE, Scores

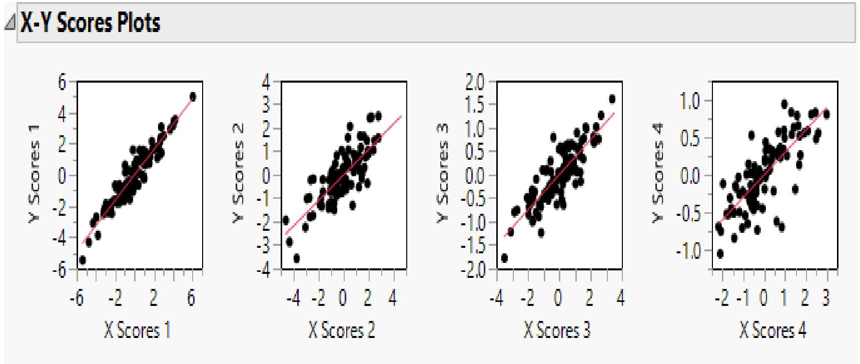
   monitoring process
3. Contribution plots

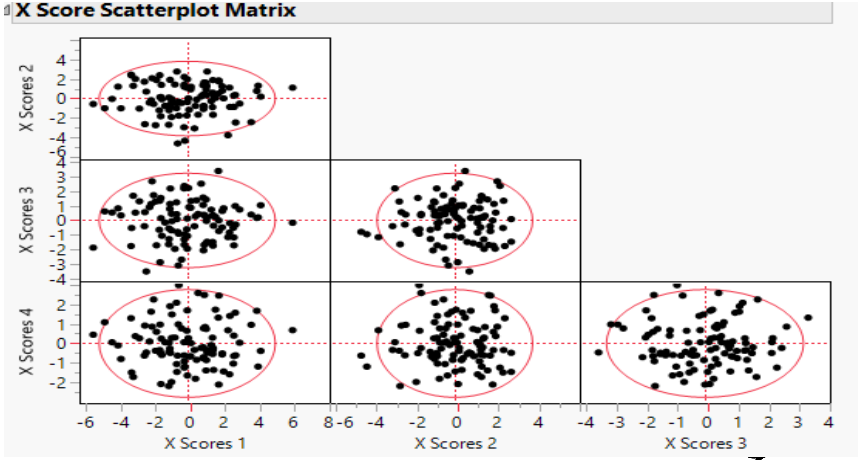
   Diagnosing process

PLS Modelling Normal Behavior

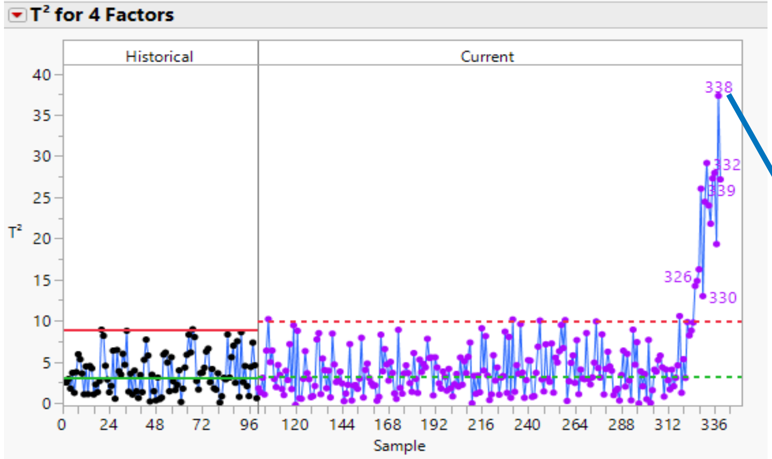
Model the normal process behavior

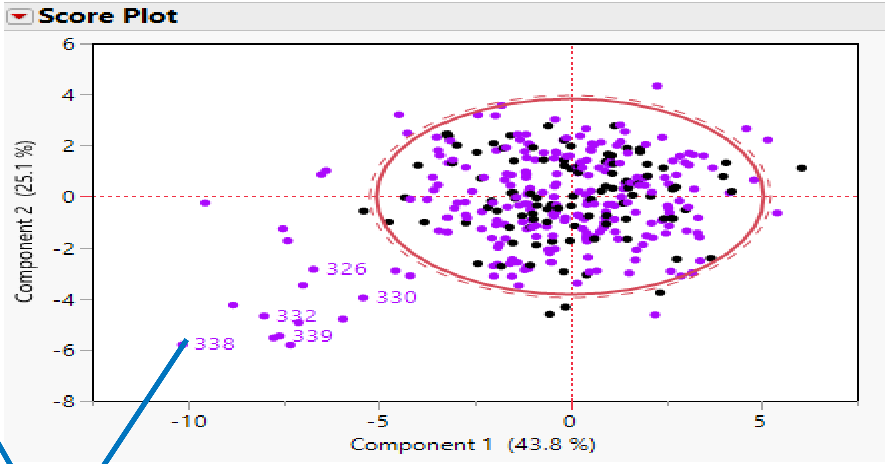
- Make a PLS on the 100 first observations
  - Four components are decided by cross validation
  - Four significant components modelling 70 % of the variability of process variables (X) and 79% of the variability of quality variables (Y)
  - Score plots show a good correlation between U and T, which means there is information in X describing Y
- Using these 100 observations, one can define areas or intervals where a process seems to be in control
  - These limits can then be used when monitoring the process





Monitoring the Process (T^2)



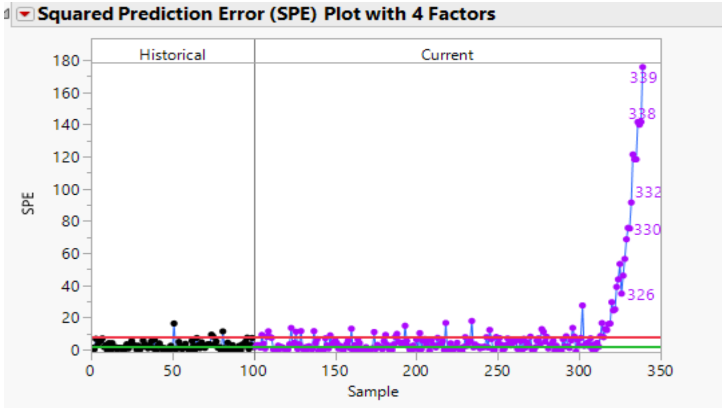


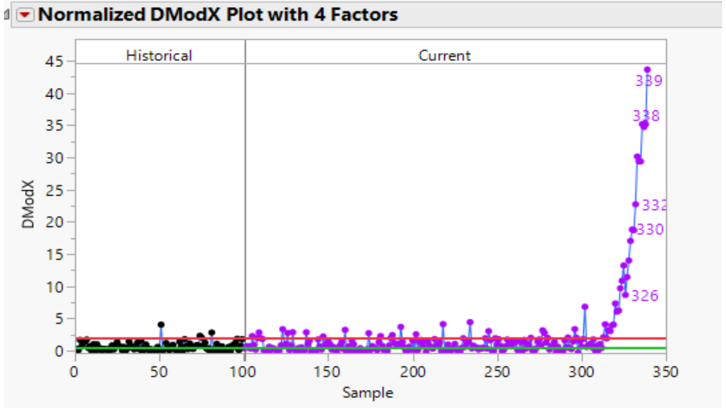
$$T_i^{\,2} = t^T \Lambda^{-1} t = \sum_{k}^{a} \left(\frac{t_k}{s_k}\right)^2 \sim \frac{a(n^2-1)}{n(n-a)} F_{a,n-a,\alpha}$$

$$T_{338}^{\;2} = \left(\frac{-10.13}{2.08}\right)^2 + \left(\frac{-5.8}{1.58}\right)^2 + \left(\frac{-0.35}{1.33}\right)^2 + \left(\frac{-0.76}{1.14}\right)^2 = 37.63$$

- In the T^2 plot we can see that most of the new observations are below the critical limits, except for observations after 325.
- Projecting the new data onto the model (t_1-t_2 plane) clearly indicates the process upset progressively move outside the acceptance region.

Monitoring the Process (SPE and DModX)





$$SPE_i = \boldsymbol{e_i}^{\boldsymbol{T}}\boldsymbol{e_i} = \sum_{j}^{p} \boldsymbol{e_{ij}^2} \sim \boldsymbol{g\chi^2_{h,\alpha}}$$

$$DModX_i = \frac{SPE_i/\boldsymbol{df_1}}{SPE/\boldsymbol{df_2}} \sim \boldsymbol{F_{h,nh,\,\alpha}}$$

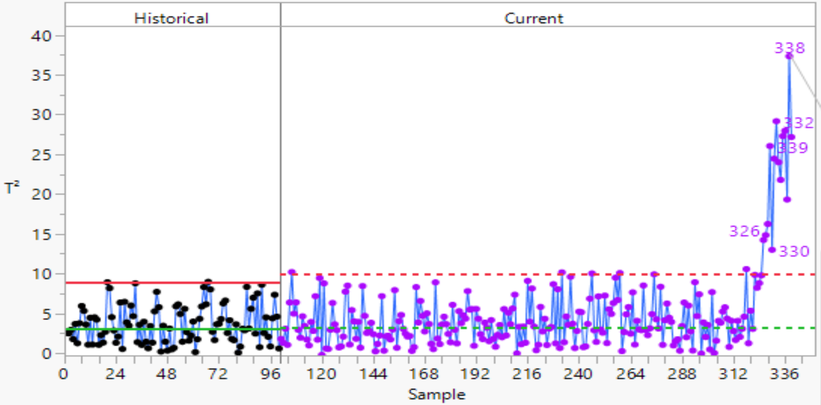
- SPE is the sum of squared prediction error and DModX is a scaled version of SPE. Both measures the distance between the observations and the model plane.

- Same conclusions can be drawn by SPE and DModX plots. Choosing which one depends on your preference.

Diagnosing the Process (T^2 Contribution Plots)

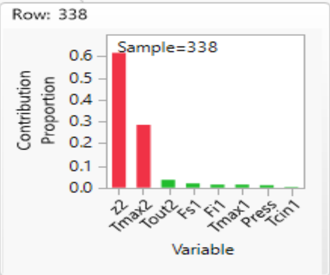
Monitor the Process

T^2 for 4 Factors



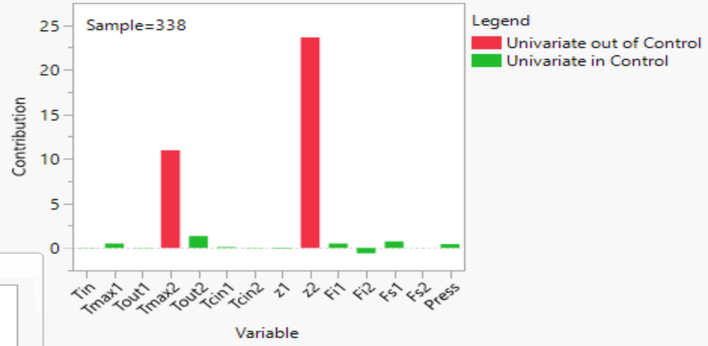
T^2 Limit Summaries

Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	3.368	9.224	1	0.05
Current	3.521	10.276	19	



Diagnose the Process

T^2 Contribution Plot for Selected Samples



$$T_i^2 = \sum_{j}^{p} Con(Ti^2, i, j)$$

$$Con(Ti^2, i, j) = \sum_{k}^{A} t_k \, s_k^{-2} p_{jk} x_{ij}$$

Identifying T^2 contributions of variables

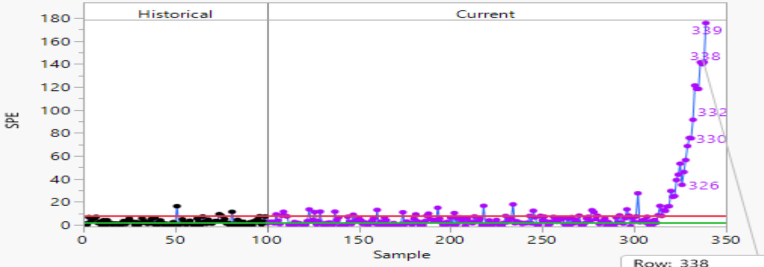
Major contributions come from Z_2 (position of the reactor where T_{max2} appears) and T_{max2} to the observation 338.

$Con(T_{338}^2, 338, Z_2) = (-10.13/4.33)*(-6.15) + (-5.8/2.5)*(-4.2) + (-0.35/1.768)*(-1.51) + (-0.76/1.3)*(1.23) = 23.64$

Diagnosing the Process (SPE Contribution Plots)

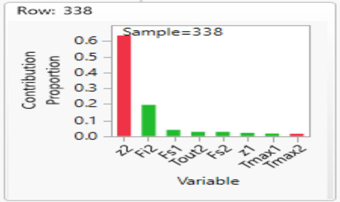
Monitor the Process

Squared Prediction Error (SPE) Plot with 4 Factors



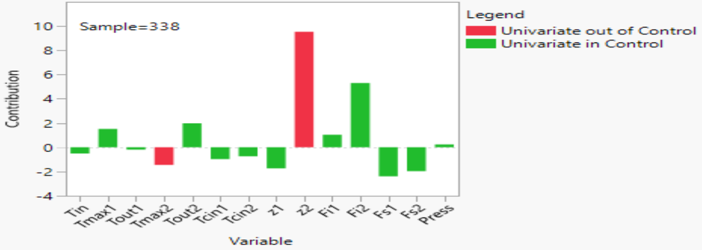
SPE Limit Summaries

Data Points	Median	UCL	Out of Control Point	Alpha Level
Historical	3.385	9.512	4	0.05
Current	3.385	9.512	49	

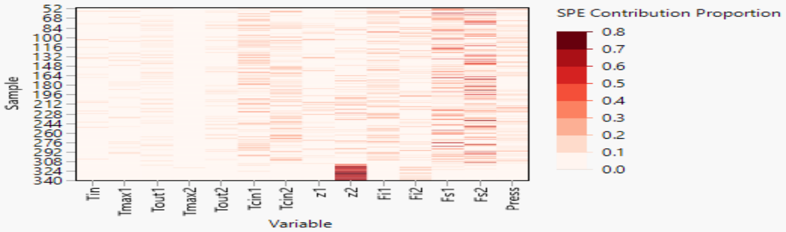


Diagnose the Process

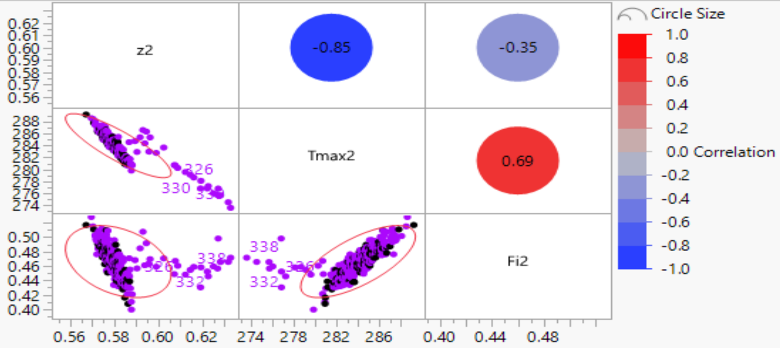
SPE Contribution Plot for Selected Samples



SPE Contribution Proportion Heat Map



Scatterplot Matrix

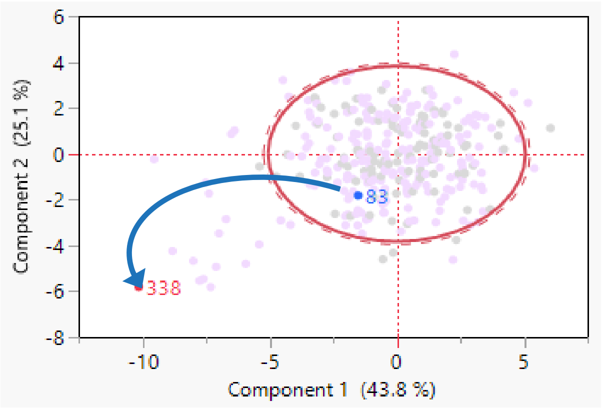


SPE contribution plots

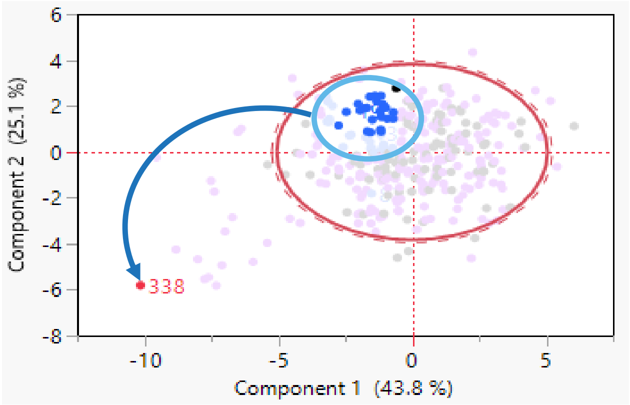
- Major contributions come from Z_2 (position of the reactor where T_{max2} appears) and F_{i2} (feed rate of the initiators to the second section).
- Hot spot position has moved further down the reactor and possibly the initiator efficiencies have dropped.
- Variable Z_2 ,T_{max2} and F_{i2} break the correlation structure.

Diagnosing the Process (Relative Score Contribution Plot)

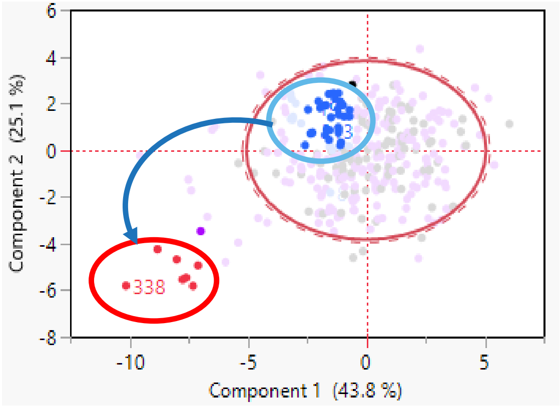
point-to-point

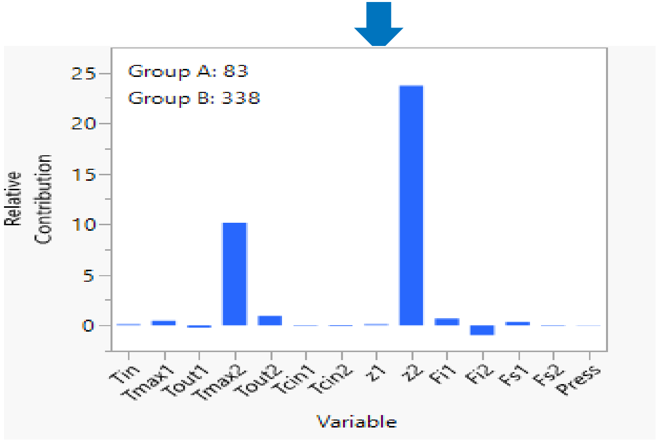


point-to-group



group-to-group



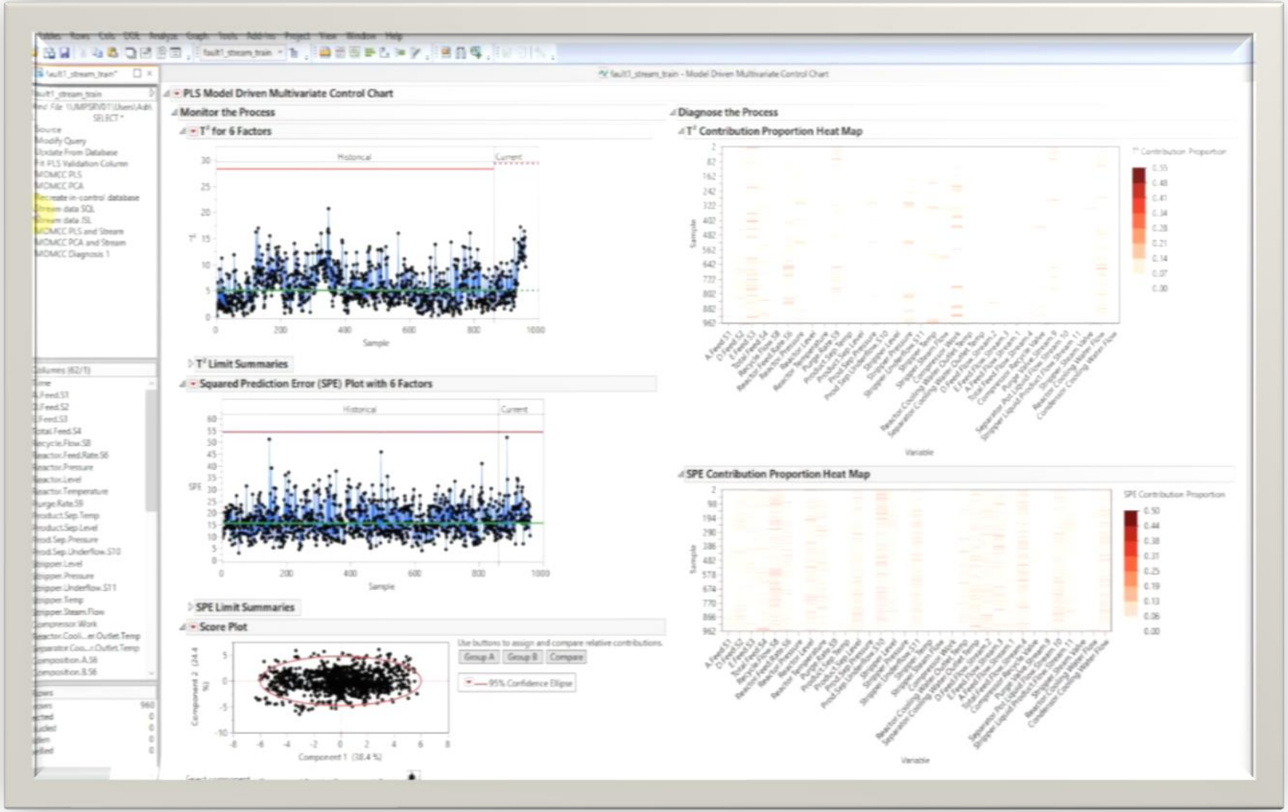


Identifying relative contributions of variables

- Major contributions to the difference between observation 83 and 338 come from Z_2 (position of the reactor where T_{max2} appears) and T_{max2}
- Hot spot position has moved further down the reactor and the hot-spot temperature T_{max2} has decreased.

Example-Tennessee Eastman Process

Process data streaming



jmppublic

jmplive

Conclusions

How the model driven multivariate control chart platform can help?

- MSPC (PCA / PLS)
  - Efficient
- Monitoring
  - Early
- Diagnosing
  - Easy
- Streaming
  - Effective

References

- Kourti, T. and MacGregor, J. F. (1995), "Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods", *Chemometrics and Intelligent Laboratory Systems*, 28, 3–21.
- **Contribution plots**: P Miller, RE Swanson, CE Heckler, "Contribution Plots: a Missing Link in Multivariate Quality Control", *Applied Mathematics and Computer Science*, *8* (4), 775-792, 1998.
- **Process monitoring**: John MacGregor and Theodora Kourti "Statistical Process Control of Multivariate Processes", *Control Engineering Practice*, **3**, p 403-414, 1995.
- Li, G., Qin, S.-Z., Ji, Y.-D., & Zhou, D.-H. "Total PLS Based Contribution Plots for Fault Diagnosis". *Acta Automatica Sinica*, *35*(6), 759–765, 2009.
- Downs, J. J., & Vogel, E. F. "A plant-wide industrial process control problem". *Computers & Chemical Engineering*, *17*(3), 245–255, 1993.

Questions?

Thank you for your attention

Jianfeng.Ding@jmp.com