

藉由JMP機器學習達成 預測品質控制/預測維護及 智慧製造..以化工(材)為例

黃崇仁 博士

黃譯賢 博士

凱律科技股份有限公司

JEN@PHOENIC-TECH.COM.TW

886-963488517

DECEMBER 28,2021

JMP作為業界頂尖的資料分析品牌，提供多樣的預測建模和資料採擷手段，並持續被Forrester評為資料科學、機器學習和預測分析領域的領導者，其系列產品廣泛地被財富500強公司、世界50強半導體企業、主要政府機構所使用

JMP – 全球先進企業的首選分析軟體



注：客戶包括但不限於此

- JMP尤其適合非統計專業背景的資料分析人員；同時又能為專業人士提供強大的分析工具
- JMP的應用非常廣泛，全球用戶已經超過30萬，主要涉及行業包括半導體和電子、醫藥、化工、食品、金融和服務、政府和教育等。JMP為全球頂尖企業客戶提供了豐富的資料分析手段與策

凱律科技股份有限公司簡介

4

- ▶ 我們是大數據分析(Big Data)、模擬、建模、預測(機器學習 Machine Learning)軟體SAS JMP台灣代理商(A partner of SAS JMP in Taiwan) ;
- ▶ 我們是加速建設Industry 4.0 & AI之基礎：『建構數據分析模型(Modeling/Big Data)』及製程、品質、實驗設計(研發)、專利問題實務解決方案提供者；
- ▶ 我們的客戶包含半導體封裝測試產業、精密機械設備產業、電子電機產業、自動化產業、記憶體製造產業、化材產業及其他。

客戶實際改善

5

提高“效率”至關重要

未使用JMP

- 研發成本高
- 研發到量產週期長
- 資料分析手段有限
- 資料分析效率低
- 良率提升慢、良率不穩

目標 =

解決方案 =

實現結果 =

使用JMP後

- 持續縮短研發週期
- 降低研發成本
- 持續改善製程與品質流程

- 建立全球統一的資料分析平臺
- 持續提高工程師與管理層的資料分析與決策能力
- 讓研發、製程、生產、庫存等關鍵領域用資料溝通

- 自上而下驅動
- 形成量化驅動決策的企業文化，而JMP正是助力文化落地的標準工具之一

製造產業常見問題

- ▶ 品質缺陷
- ▶ 精度缺失
- ▶ 設備故障
- ▶ 整體運轉效率損失

製造業為什麼需要機器學習？

- ▶ 在資料無處不在、資料量爆炸式增長的今天，如何高效的利用資料、選擇適合的分析方法成為了關鍵，機器學習正是在這樣的背景下走入工程師的視野
- ▶ **機器學習**，是一種可以基於樣本資料自動化構建數學模型，以便能夠基於相同產品投入新批次生產時用來預測產出結果是否在規格內的分析方法

機器學習預測建模流程

8

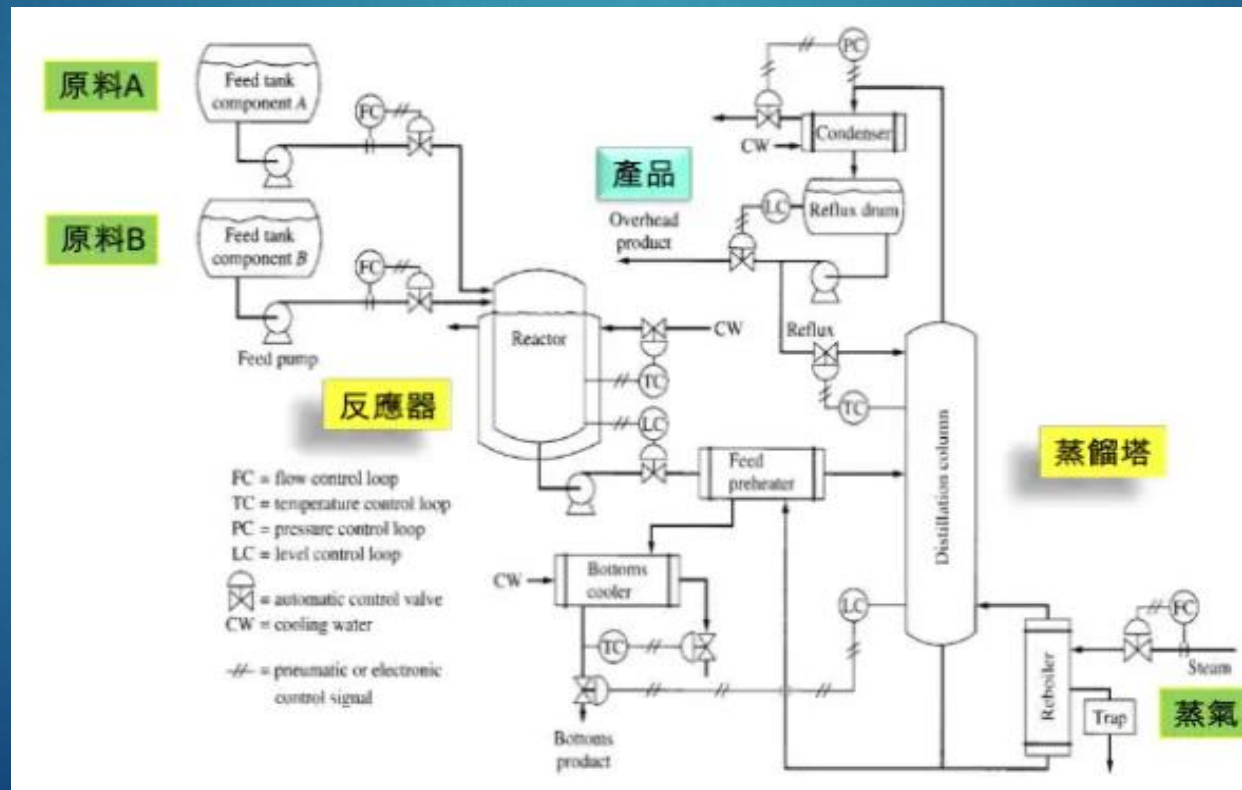


案例背景

9

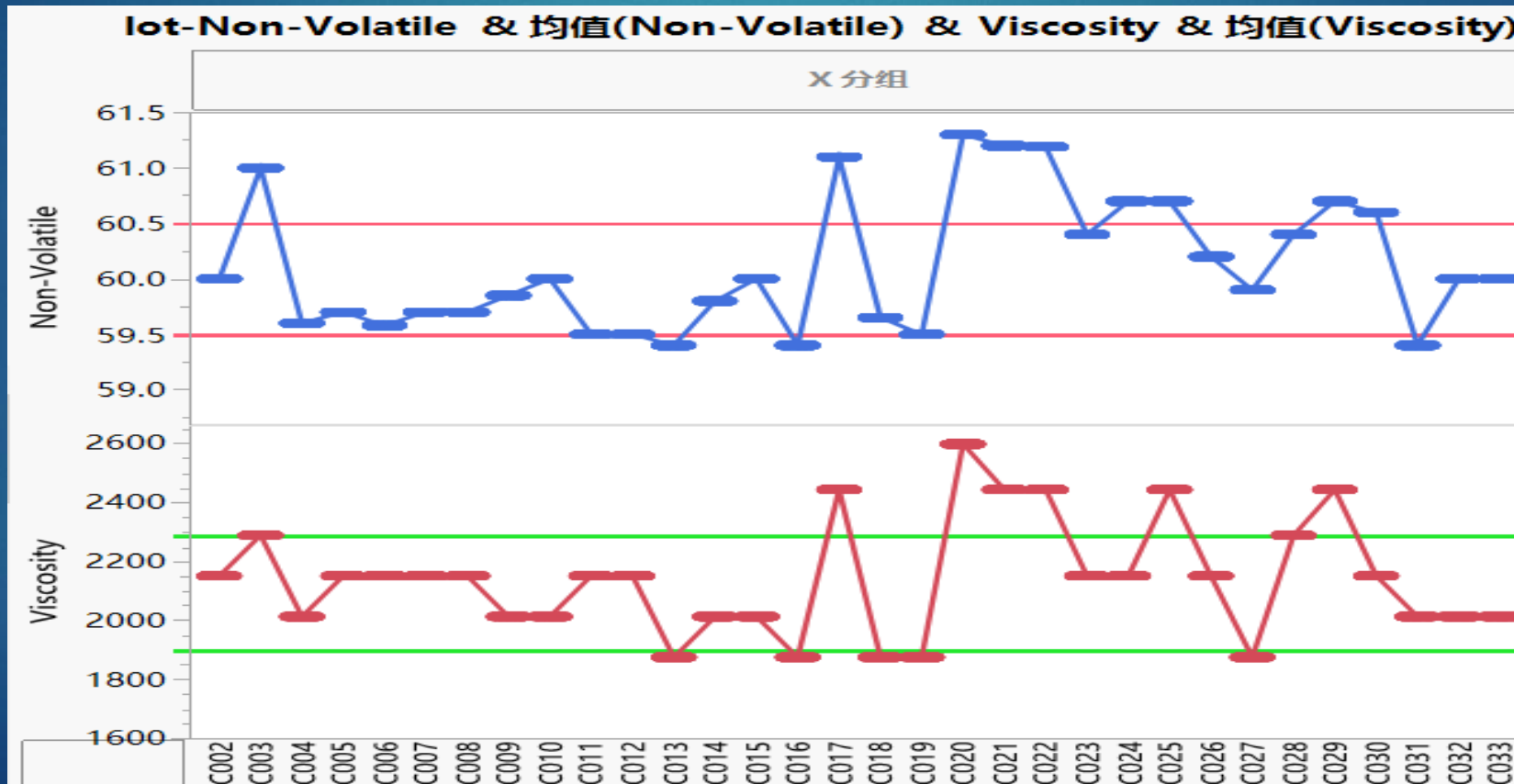
化工(材)生產產品時，每批次生產時會有反應過程中**每分鐘**紀錄的生產資料(X)及最後反應完後採樣的生產數據(Non-Volatile, Viscosity)(Y)，資料量龐大，而且**異常複雜**；
在製造過程的最後 **每批次只有一組產出結果數據**，需要知道那個製程參數(X) 是**關鍵影響**最後產出結果(是否在USL及LSL內)

Dr. C.J Huang Edited, copy rights reserved
2022/1/10



Step 1 定義業務問題

工程師在某化工(材)產品之控制系統分析(Process Control)觀測到多批製程之產出規格(USL及LSL)有差異 (在某些化工產品上的良率損失高達20%)，推測該損失與**某些製程參數**效應有關，因此，一旦確定了**某些製程參數**，就很容易找到的**關鍵過程參數**，從而形成**糾正和預防措施**。



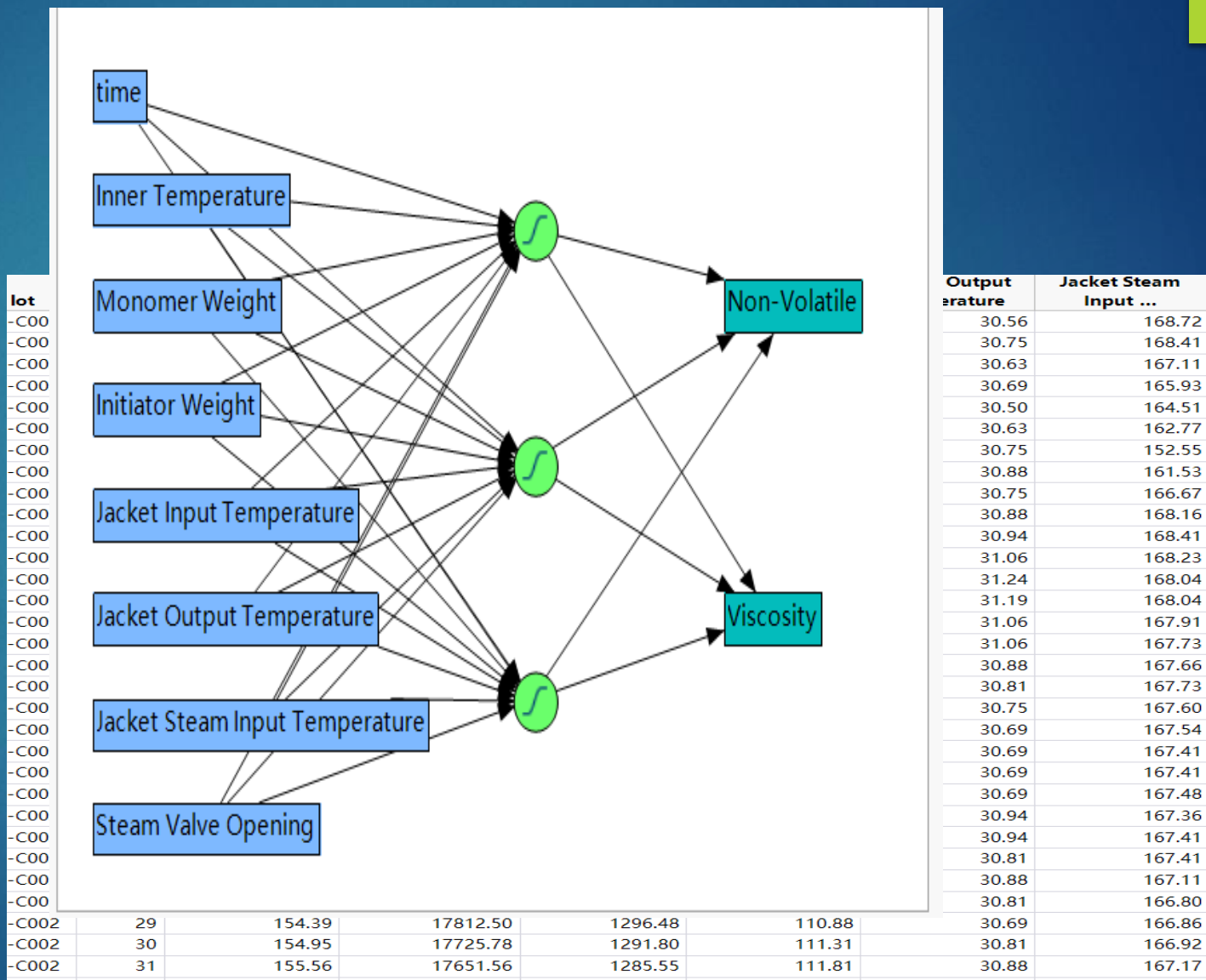
Step 2 資料獲取和準備

將用於良率分析的資料進行關聯；比如，按 Lot ID + Reaction Time (化學反應時間)，Parameter，Equipment，甚等 Inline 資料；

lot	time	Inner Temperature	Monomer Weight	Initiator Weight	Jacket Input Temperature	Jacket Output Temperature	Jacket Steam Input ...	Steam Valve Opening	N
-C002	1	158.79	19925.78	1440.43	106.30	30.56	168.72	110	
-C002	2	158.91	19907.03	1434.38	106.30	30.75	168.41	0	
-C002	3	159.16	19913.28	1431.25	107.10	30.63	167.11	0	
-C002	4	159.16	19900.78	1431.25	107.54	30.69	165.93	0	
-C002	5	158.98	19832.81	1428.13	109.52	30.50	164.51	0	
-C002	6	157.80	19764.84	1425.00	109.89	30.63	162.77	0	
-C002	7	155.88	19678.13	1425.00	108.65	30.75	152.55	100	
-C002	8	154.33	19596.88	1415.82	106.91	30.88	161.53	100	
-C002	9	153.40	19498.44	1406.45	108.34	30.75	166.67	100	
-C002	10	152.84	19423.44	1400.20	109.02	30.88	168.16	100	
-C002	11	152.47	19330.47	1397.07	109.27	30.94	168.41	100	
-C002	12	152.03	19243.75	1391.02	110.27	31.06	168.23	100	
-C002	13	151.84	19182.03	1384.77	111.31	31.24	168.04	100	
-C002	14	151.66	19070.31	1378.52	111.88	31.19	168.04	100	
-C002	15	151.54	19008.59	1375.39	112.00	31.06	167.91	100	
-C002	16	151.48	18921.88	1369.34	112.18	31.06	167.73	100	
-C002	17	151.35	18835.16	1363.09	112.00	30.88	167.66	100	
-C002	18	151.35	18748.44	1356.84	111.38	30.81	167.73	100	
-C002	19	151.35	18661.72	1350.59	110.76	30.75	167.60	100	
-C002	20	151.48	18568.75	1347.66	110.38	30.69	167.54	100	
-C002	21	151.60	18494.53	1339.84	109.77	30.69	167.41	100	
-C002	22	151.84	18407.81	1335.16	109.89	30.69	167.41	100	
-C002	23	152.09	18314.84	1332.03	110.33	30.69	167.48	100	
-C002	24	152.41	18246.88	1325.98	110.76	30.94	167.36	100	
-C002	25	152.71	18160.16	1319.73	110.95	30.94	167.41	100	
-C002	26	153.02	18073.44	1311.91	110.88	30.81	167.41	100	
-C002	27	153.46	17998.44	1308.79	110.51	30.88	167.11	100	
-C002	28	154.02	17893.75	1302.73	110.63	30.81	166.80	100	
-C002	29	154.39	17812.50	1296.48	110.88	30.69	166.86	100	
-C002	30	154.95	17725.78	1291.80	111.31	30.81	166.92	100	
-C002	31	155.56	17651.56	1285.55	111.81	30.88	167.17	100	

Step 4 預測建模方法 選擇

工程師最終發現基於
決策樹的集成演算法
**神經網絡提升(Neural
network
boosted)** 等機器學習
演算法更適合應對此
類問題



Step 5 模型診斷及結果分析

從模型的輸出報表，可以看到：
訓練集獲得了較高的 R2，驗證集和測試集的 R2 損失較小，並且兩者的偏差較小，因此，**所建立的模型可以很好地解釋過程變異**
此外，通過查看對**神經網絡**提有**非零**貢獻的列變數，識別可能影響 Y 的因數，**而根因將是那些貢獻最大的因數，因為特定變數與回應的相關性越大，被選擇的頻率就越高**

神经
验证列: 验证

模型启动

模型: N1anH(3)NBBoost(6)

训练		验证	
Non-Volatile		Non-Volatile	
测量	值	测量	值
R 方	1	R 方	1
RASE	8.396e-10	RASE	8.5e-10
绝对偏差的均值	5.78e-10	绝对偏差的均值	6.052e-10
-对数似然	-110836.2	-对数似然	-73837.8
误差平方和	4.011e-15	误差平方和	2.74e-15
频数总和	5690	频数总和	3793

Viscosity		Viscosity	
测量	值	测量	值
R 方	0.8547444	R 方	0.8475665
RASE	74.406594	RASE	76.065596
绝对偏差的均值	63.496689	绝对偏差的均值	65.658982
-对数似然	32595.069	-对数似然	21811.778
误差平方和	31501782	误差平方和	21946203
频数总和	5690	频数总和	3793

	广义 R 方	-对数似然
训练	1.0000	-78241.14
验证	1.0000	-52026.02

拟合组

"Non-Volatile" 的 Bootstrap 森林法

总体统计量

单个树	RASE
平均	0.3 / 88661

R 方	RASE	数目
0.882	0.2028759	9483

按树汇总

列贡献

项	拆分数	平方和	对应部分
Jacket Input Temperature	9520	472.603288	0.2703
Jacket Output Temperature	6200	425.920229	0.2436
Initiator Weight	6009	239.383024	0.1369
Jacket Steam Input Temperature	7160	219.533526	0.1255
Monomer Weight	5569	155.478478	0.0889
time	5840	112.820475	0.0645
Inner Temperature	5078	98.4868423	0.0563
Steam Valve Opening	1681	24.4850885	0.0140

Step 6 根因識別

工程師將列貢獻最大的變數提取出來後發現：

- 列貢獻大 (本例選取Top4) 的預測變數之間可能存在共線性，因此，可以結合多元相關分析(多元控制圖)或聚類變量和工程經驗進一步甄別；
- 對於本例，工程師最終 (Inner temperature & Steam valve opening 為一群, Monomer weight & Initiator Weight 為一群) 識別為最重要的影響因數

聚類成員				
聚類	成員	与聚類自身的 R 方	与最邻近聚類的 R 方	1-R 方比
1	Jacket Steam Input Temperature	0.905	0.153	0.112
1	Initiator Weight	0.904	0.357	0.149
1	Monomer Weight	0.889	0.365	0.175
1	Jacket Input Temperature	0.563	0.069	0.469

主成分: 基于协方差										
特征值	百分比	30	40	50	60	80	累积百分比	卡方	自由度	概率>卡方
44174360	99.872						99.872	883992	54.000	<.0001*
37470.66	0.085						99.957	290771	44.000	<.0001*
8989.375	0.020						99.977	228572	35.000	<.0001*
8099.005	0.018						99.995	205791	27.000	<.0001*
1169.645	0.003						99.998	133430	20.000	<.0001*
382.9809	0.001						99.999	112086	14.000	<.0001*
277.9798	0.001						99.999	106048	9.000	<.0001*
233.6608	0.001						100.000	98702.0	5.000	<.0001*
1.9085	0.000						100.000	18743.7	2.000	<.0001*
0.0870	0.000						100.000	0.000	0.000	.

特征向量	time	Inner Temperature	Monomer Weight	Initiator Weight	Jacket Input Temperature	Jacket Output Temperature	Jacket Steam Input Temperature	Steam Valve Opening	Non-Volatile	Viscosity
time	-0.00000	0.00001	0.00041	-0.00082	0.00247	0.03536	0.01289	-0.04440	0.00783	0.00127
Inner Temperature	-0.00000	-0.00000	0.00002	0.00000	-0.00005	-0.00250	0.00192	-0.00157	0.72241	-0.00696
Monomer Weight	0.00015	-0.00001	0.00037	-0.00061	0.00004	0.00032	0.00019	-0.00032	0.00014	0.00003
Initiator Weight	0.00001	0.00007	-0.00559	0.00932	-0.00118	0.00147	0.00037	-0.00628	-0.00040	0.00021
Jacket Input Temperature	0.00000	0.00005	0.00043	0.00080	0.02866	-0.00649	-0.00706	-0.00329	-0.00149	-0.00046
Jacket Output Temperature	-0.00000	-0.00000	0.00889	0.00590	-0.00189	0.00159	0.00223	0.00170	-0.00057	0.00471
Jacket Steam Input Temperature	0.00000	0.00006	-0.00067	0.00024	0.00476	0.02159	0.03697	0.04190	0.01217	-0.01110
Steam Valve Opening	0.00000	0.00001	-0.00011	0.00021	0.00002	0.02901	-0.04478	0.02229	0.04346	-0.00251
Non-Volatile	0.00000	0.00001	-0.00001	-0.00001	0.00002	0.00007	0.00008	0.00016	0.00155	3.39107
Viscosity	0.00000	0.00516	0.00008	-0.00013	-0.00029	-0.00033	-0.00027	-0.00035	-0.00015	-0.00842

注意: 对特征向量执行了除以特征值平方根的运算。

解釋了變數中大約 99.872% 的總變異

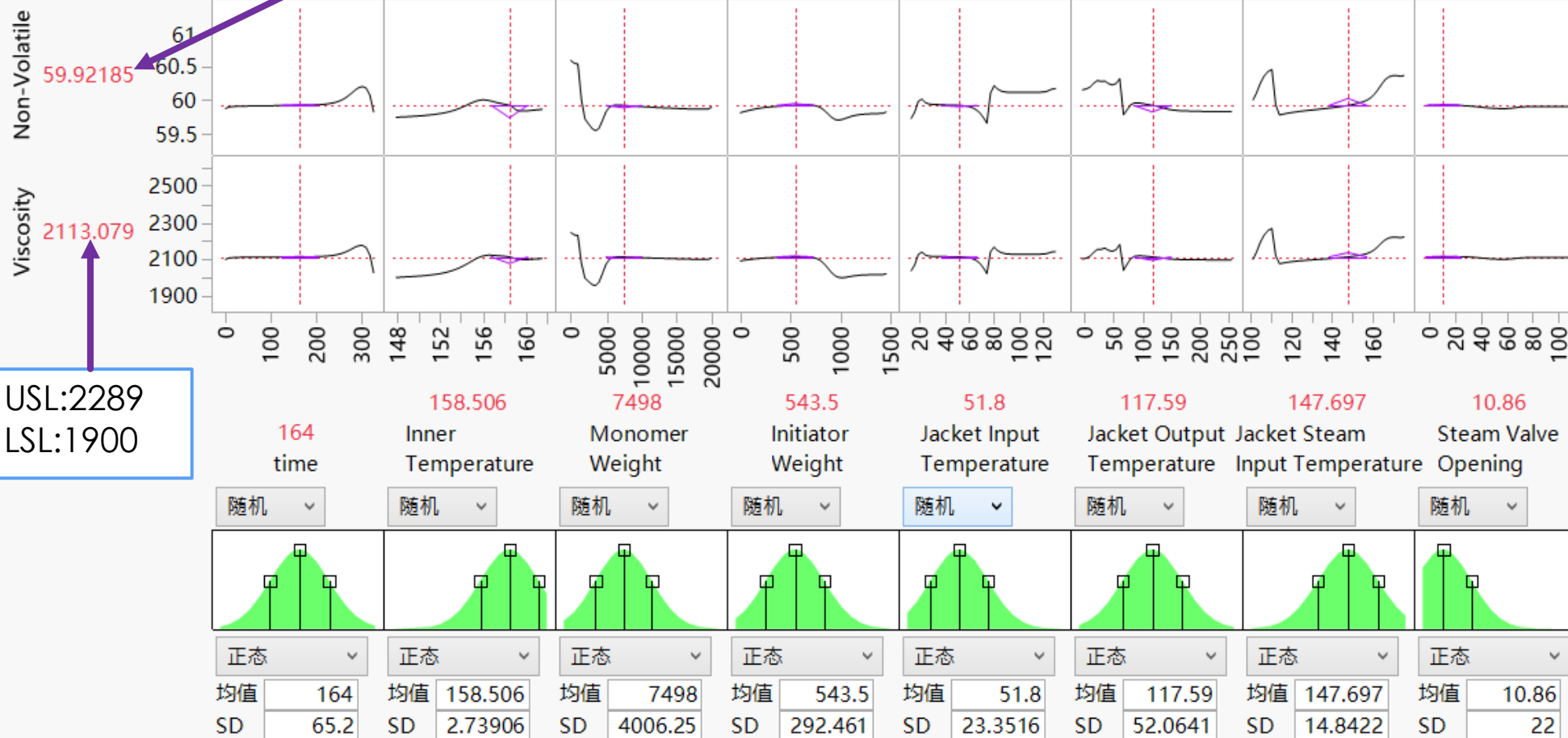
Step 7 結果驗證

- ▶ 工程師通過進一步追溯變異源，發現 **Steam Valve Opening, Inner Temperature** 和某關鍵工藝設備之間存在明顯的相關性，Tukey-Kramer 多重比較確認了該差異的顯著性；
- ▶ 因此，工程師可以展開更進一步的排查設備問題，找到導致過程輸出變異和波動的原因，從而形成有針對性的預防措施和改進措施。

Step 8.1 預測建模..神經網路..提升

USL:60.5 LSL:59.5

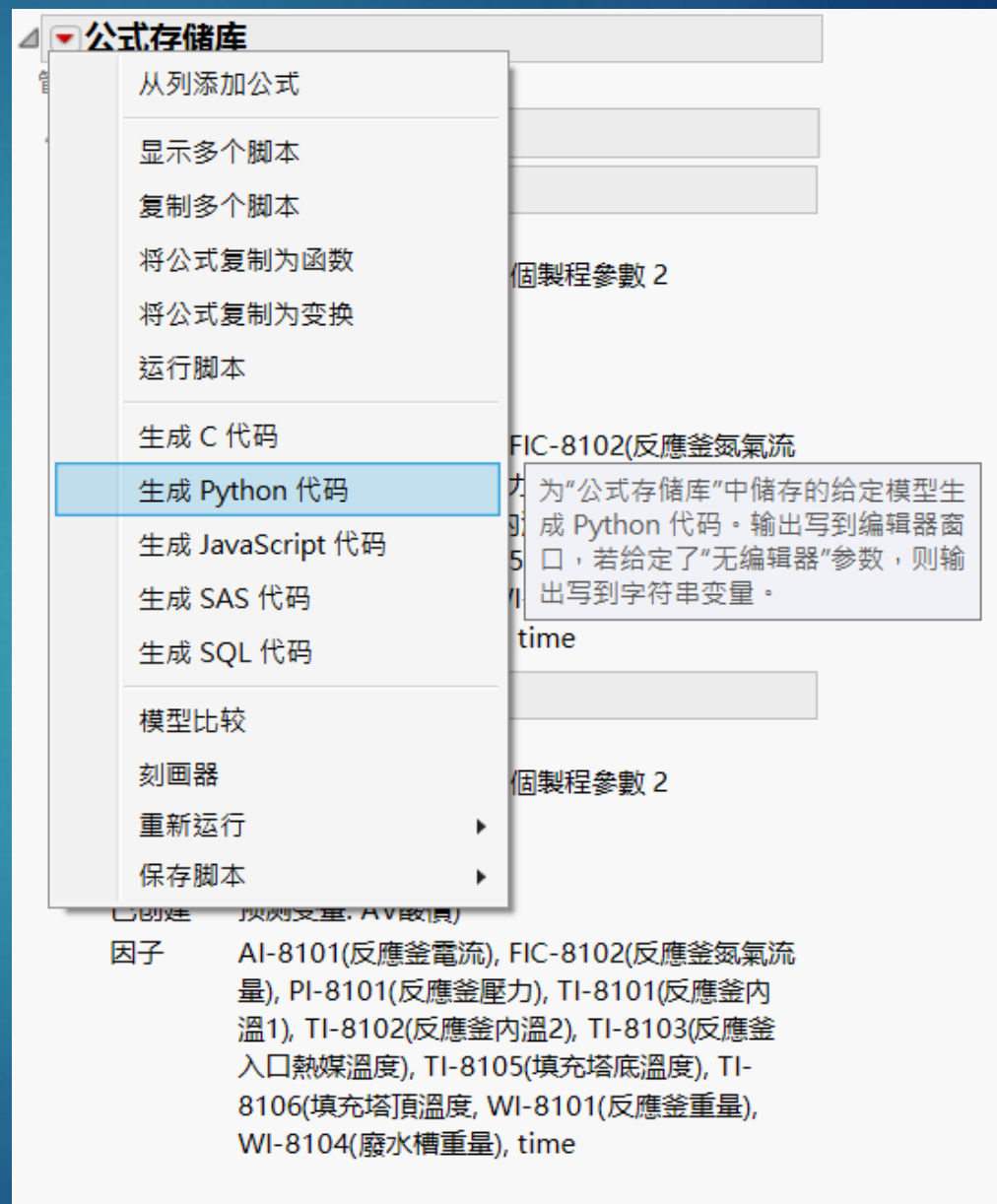
预测刻画器



USL:2289
LSL:1900

Step 8.2 機器學習

將預測模型運算公式轉換為
與機械溝通語言，例如 C
Code, Python Code, Java
Script Code, SQL Code 或
SAS Code讓機器學習



模型部署及應用

- ▶ 越來越多的高端製造企業期望借助不同的工具和方法搭建大資料分析平臺，打造智慧工廠，因此，我們將**類似本案例所獲得模型以通用代碼形式部署於企業級的大資料分析平臺**，這是實現**預測性品質控制和預測性維護**的有效手段之一，也是被證明行之有效的科學方法

敬請指教

黃崇仁 博士
黃譯賢 博士

凱律科技股份有限公司

Phoenix-iptech. Corp.

jen@phoenix-tech.com.tw

www.phoenix-
tech.com.tw

886-7-2851008

