

The team project THETA

A team for helping the industry +AI



Theta Team 專案執行成員

背景與主要專長



游為翔

- 國立台灣大學 碩士
- 中研院資訊所 資料科學家
- 講師: R 語言資料分析, 深度學習實作
- 專長: 神經科學, 機器學習/深度學習, 統計分析, 專案管理與團隊溝通



楊証琨

- 國立台灣大學 碩士
- 中研院資訊所 資料科學家
- 講師: 網路爬蟲實作, 深度學習實作
- 專長: 智慧運輸系統, 數據分析, 機器學習/深度學習, 專案管理與計畫控管



周俊川

- 國立政治大學 學士
- 中研院資訊所 資料科學家
- 專長: 統計分析, 機器學習/深度學習, 風險控管, 參數調控, 時序型資料分析與預測

重要專案經驗

- 深度卷積神經網路於印刷電路銅箔基板製造業之自動化缺陷分類與檢測
- 機器學習於機械參數調控最佳化研究與應用
- 網路書商大數據分析與洞察報告
- 電子交易票證消費之市場分析與使用率預測

- 生質能源之新陳代謝網路參數最佳化
- 深度卷積神經網路於印刷電路板之缺陷偵測
- 社交軟體詐騙帳號分析與預測
- 氣候變遷影響交通運具使用量之分析

- 遞迴式神經網路於球賽事件與博奕市場關係之預測研究
- 深度學習於染料配色之應用
- 文字探勘用於台北市政府陳情系統之滿意度分析

Theta Team 專案執行成員

背景與主要專長



杜靖凱

- 國立政治大學 學士
- 中研院資訊所 資料科學家
- 專長: 統計分析, 機器學習/深度學習, 影像處理與物體檢測, 圖形化使用者介面開發



丁新一

- 國立東華大學
- 中研院資訊所 研究助理
- 專長: 網頁前後端語言, 網路爬蟲, 影像處理, 機器學習/深度學習



林聖傑

- 國立台灣大學 碩士
- 中研院資訊所 研究助理
- 專長: 資料分析, 機器學習, 深度學習於時間序列資料分析

重要專案經驗

- 深度卷積神經網路於埋葬蟲檢測與軌跡追蹤
- 自動化圖像部位分割於蟲體之應用
- 深度學習於化學製程之產品品質預測
- 博弈行為研究 - 以關聯式規則分析玩家入場時機點

- 深度卷積神經網路於印刷電路板之缺陷偵測
- 深度學習於空壓機跳機預警分類
- 公務出國報告查詢統計網之前後端建置、資料攫取與文本分析
- 遠見雜誌專題報導 – 農委會數據攫取與整理

- CMOS image sensor 特性分析
- 電信客戶流失預測
- 深度學習於化學製程之產品品質預測

Theta Team 專案執行成員

背景與主要專長



鄭凱翔

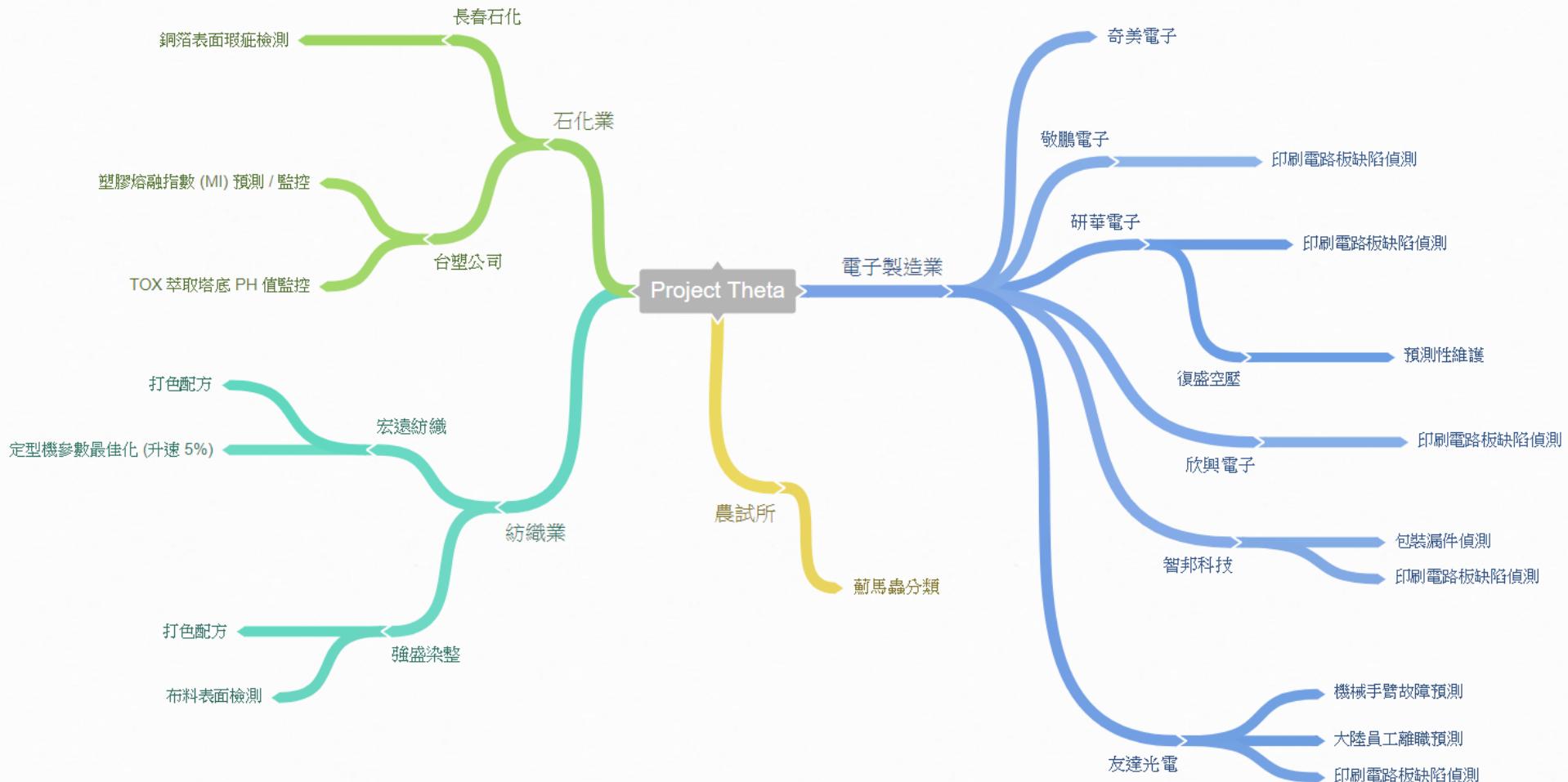
- 國立台灣大學 碩士
- 中研院資訊所 資料科學家
- 專長:機器學習 / 深度學習於參數調控,
生成式對抗網路

重要專案經驗

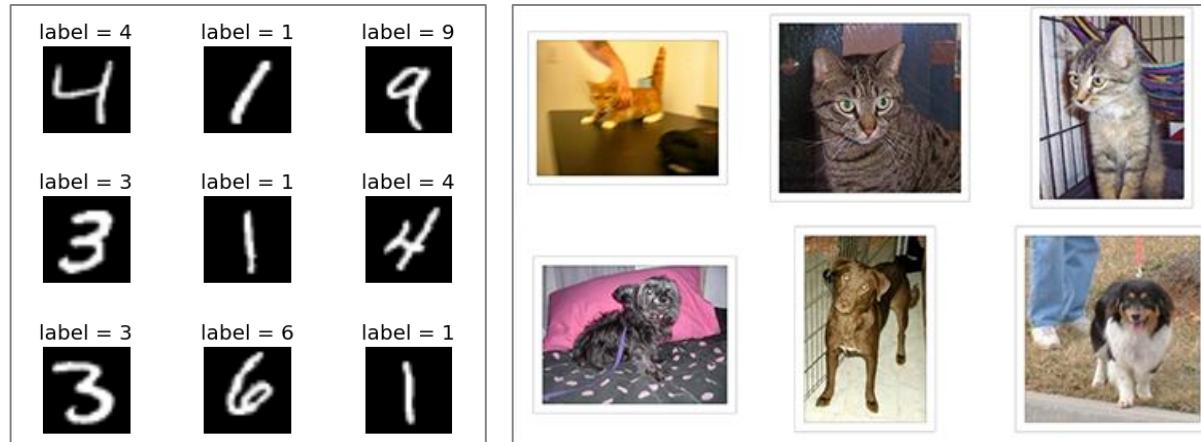
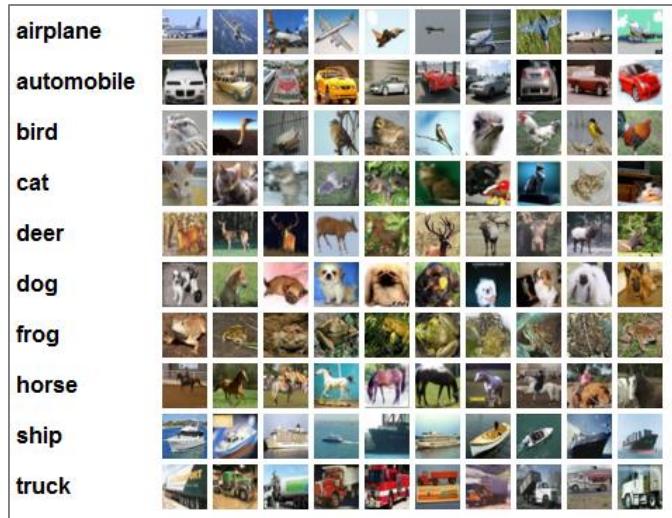
- 生成對抗網路於冷凍電子顯微鏡影像重建
- 深度學習於化學製程之產品品質預測

Proof of Concept (POC)

與數間國內製造業者共同定義挑戰



網路上的資源很多，教學也很多 直接套不難吧？

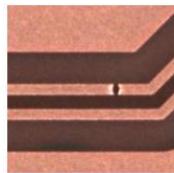


Some interesting Cases

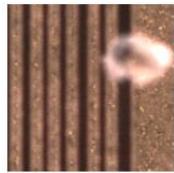
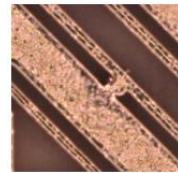
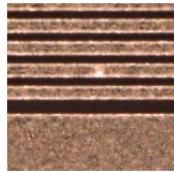
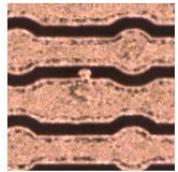
Some interesting Cases

缺陷分類問題

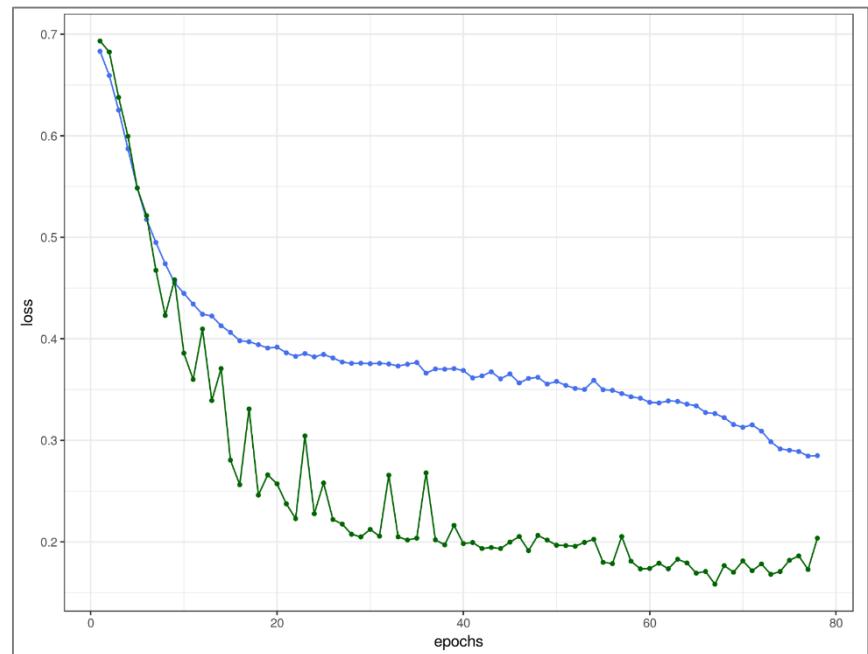
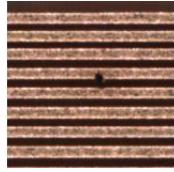
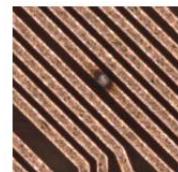
銅箔基材缺陷分類



Fail



Pass

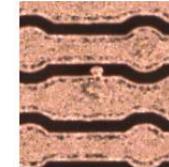


缺陷分類在現實世界中的問題

原本人工標記的 label 真的是乾淨的？

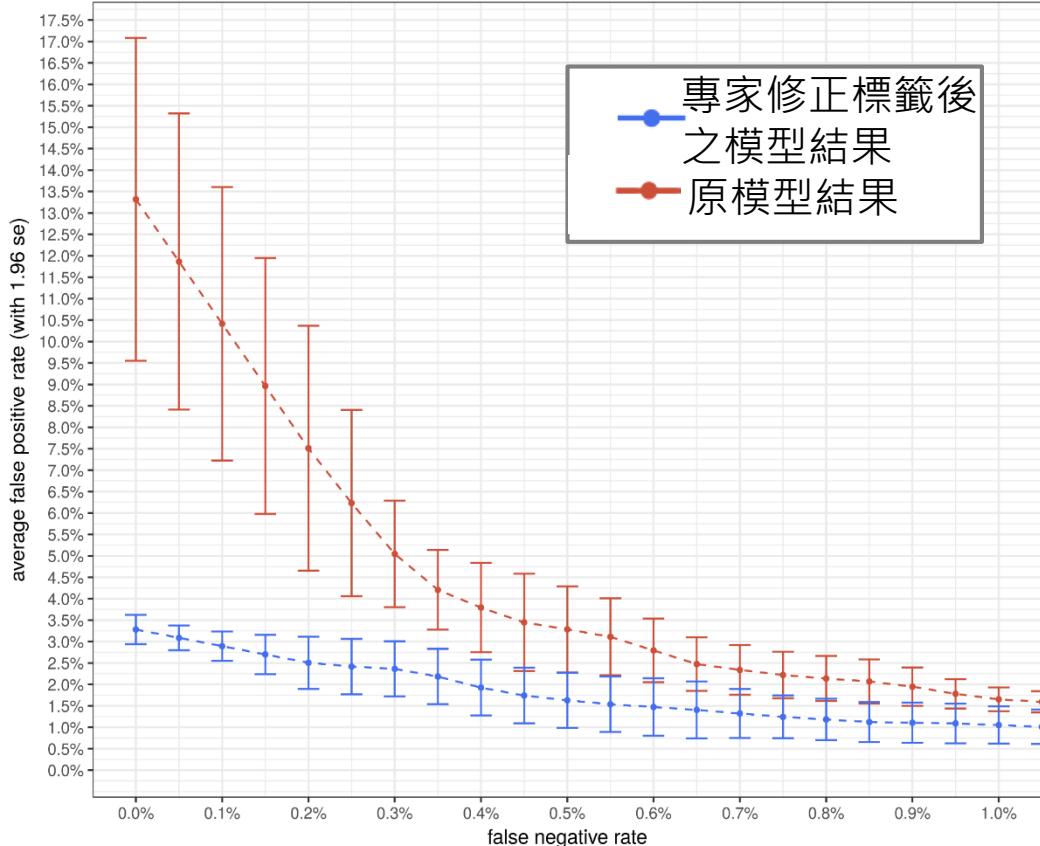
- 專家複判 Web UI

下圖是否為異常圖片 (Yes / No)



No

Average false-positive rate under fixed false-negative rate



建 Yes

↑ 或 W 鍵前一題

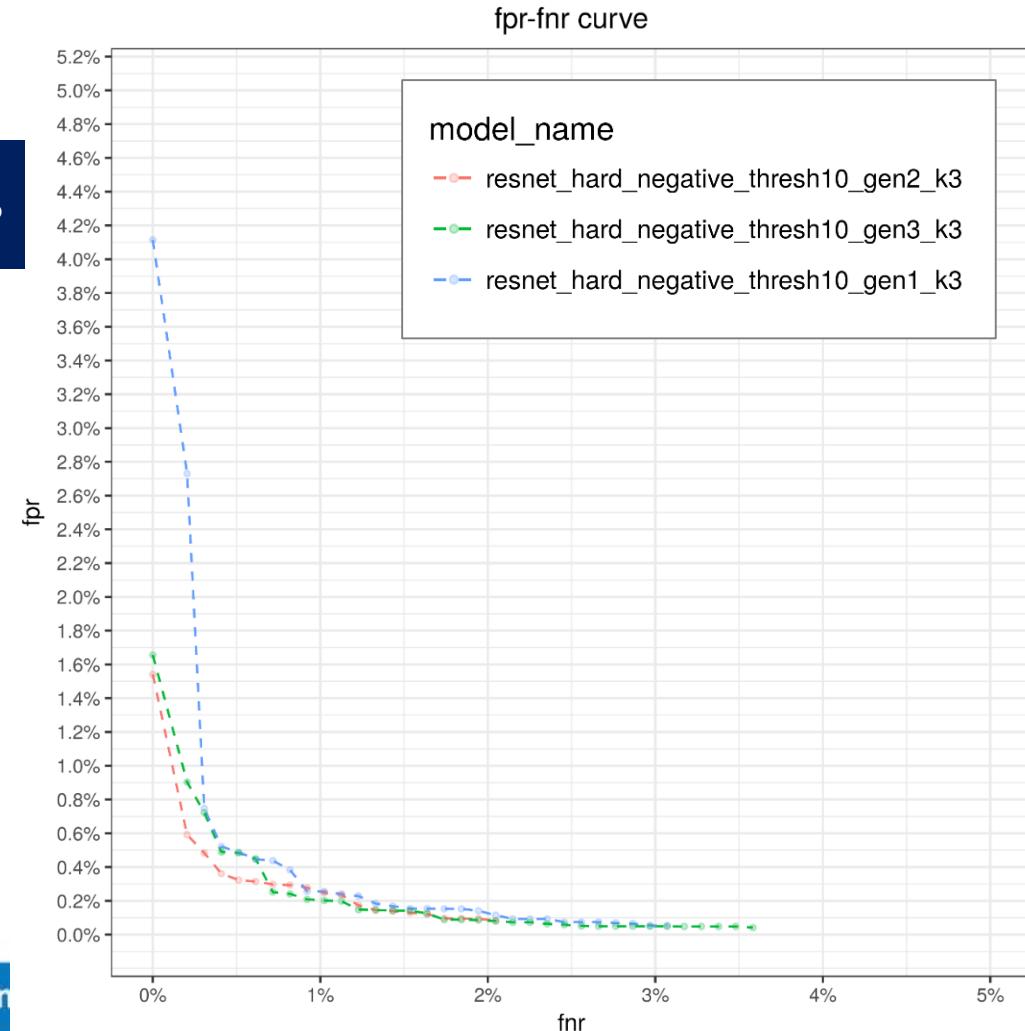
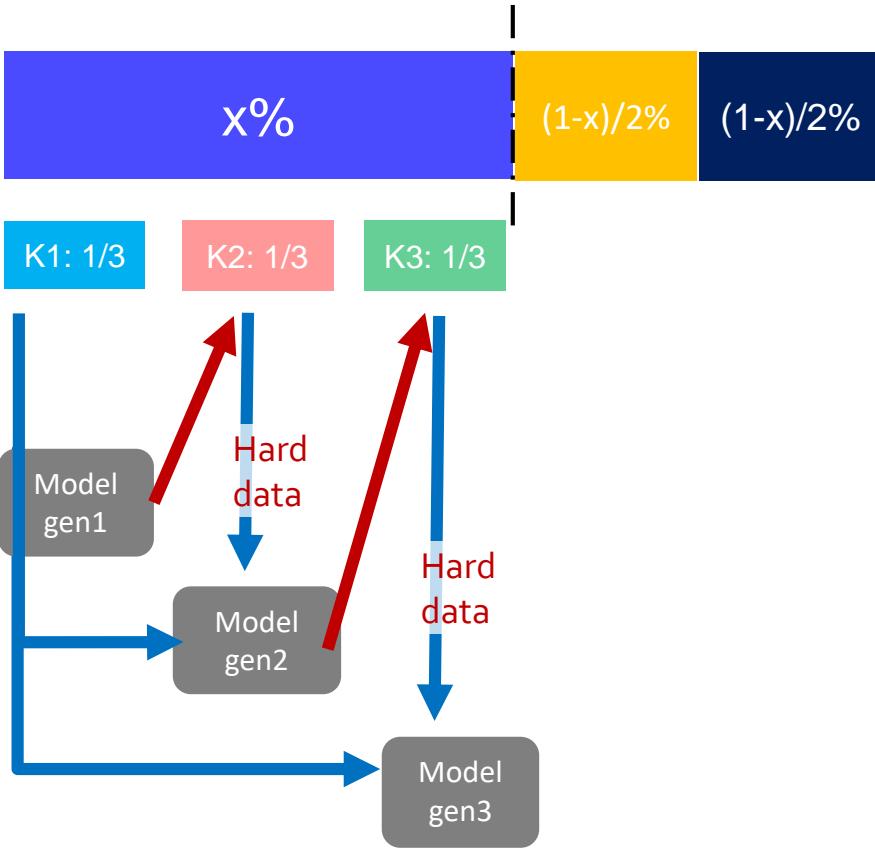
→ 或 D 鍵 No

↓ 或 Enter 或 S 鍵下一題

如何再將表現推向極致?

訓練一次不好, 你有試過訓練兩次嗎?

- Data boosting



視覺檢測：金屬原物料表面瑕疵分類效益評估

傳統
人力
目檢

產線數量：23 條，共配置 4 位後端目檢人員

自動視覺檢測系統 (Automatic Optical Inspection, AOI) 設備每小時影像輸出量：為配合人力負荷，控制在每日 120 萬張

(極限為每條產線 2 萬張/小時 = 1104 萬/日)

準確率：leakage rate 約為 1 – 3 %，overkill rate 約為 10 – 20%

硬體設備：中高階工作站級電腦 + NVIDIA GPU：約 30 萬/套

軟體：開源軟體 + 高度調校之深度學習模型

輸出量：每分鐘 1 萬張影像，每日輸出可達 1440 萬張影像，超越總產出

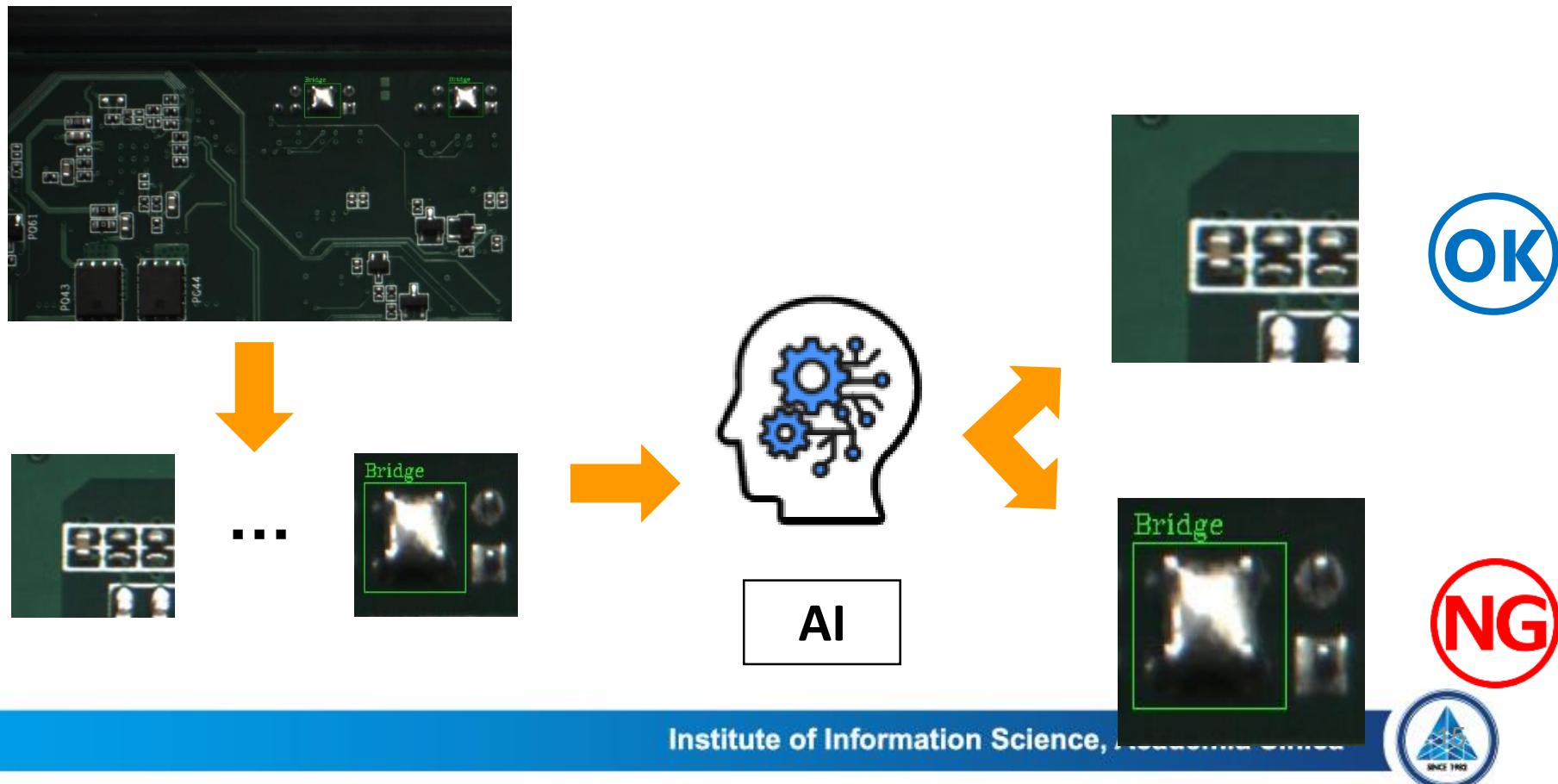
準確率：在 leakage rate 控制在 0.1% 狀況下，overkill rate 約為 1.5%

深度
學習
系統

電路板上的缺陷偵測與分類

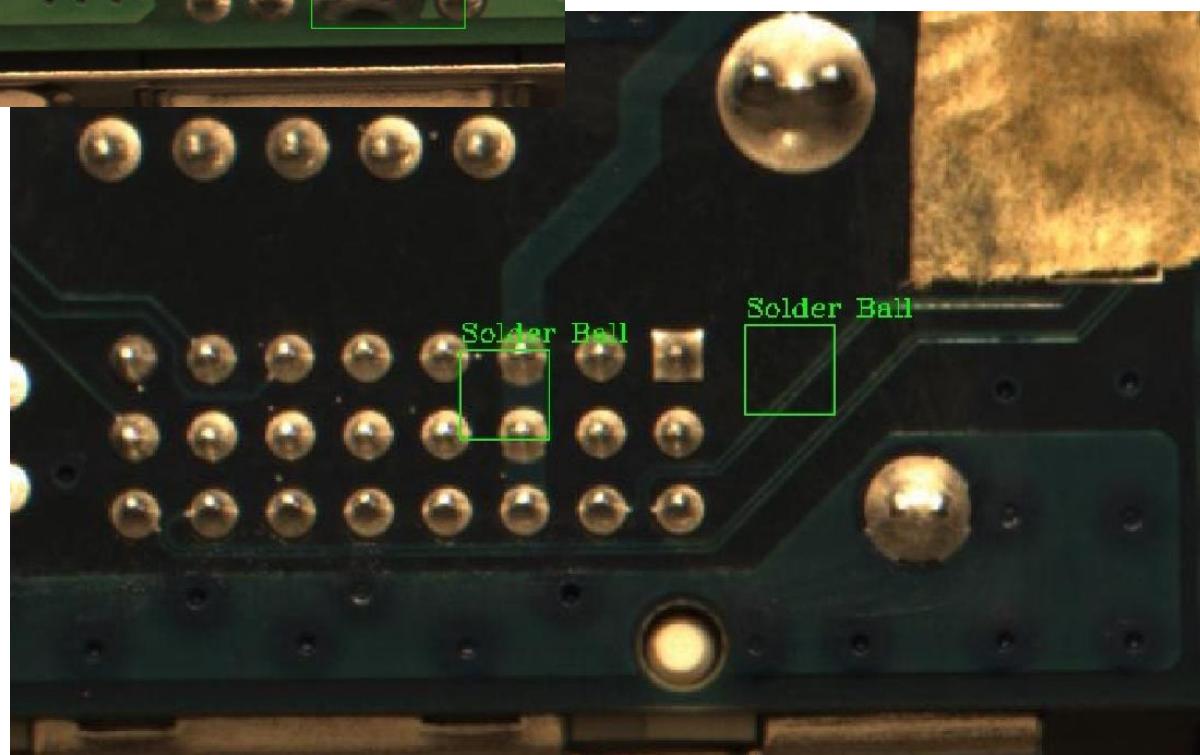
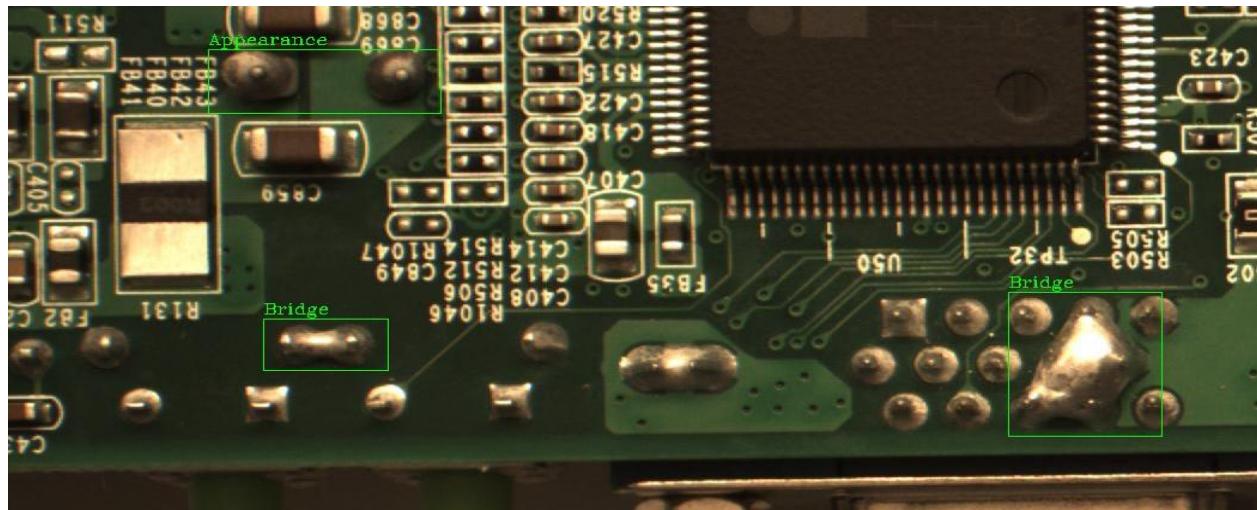
AOI: 插件、焊錫等缺陷偵測

- 生產板進入 AOI 後，透過 Deep learning module 偵測缺陷 (NG) 位置
- Transfer learning: Residual neural network



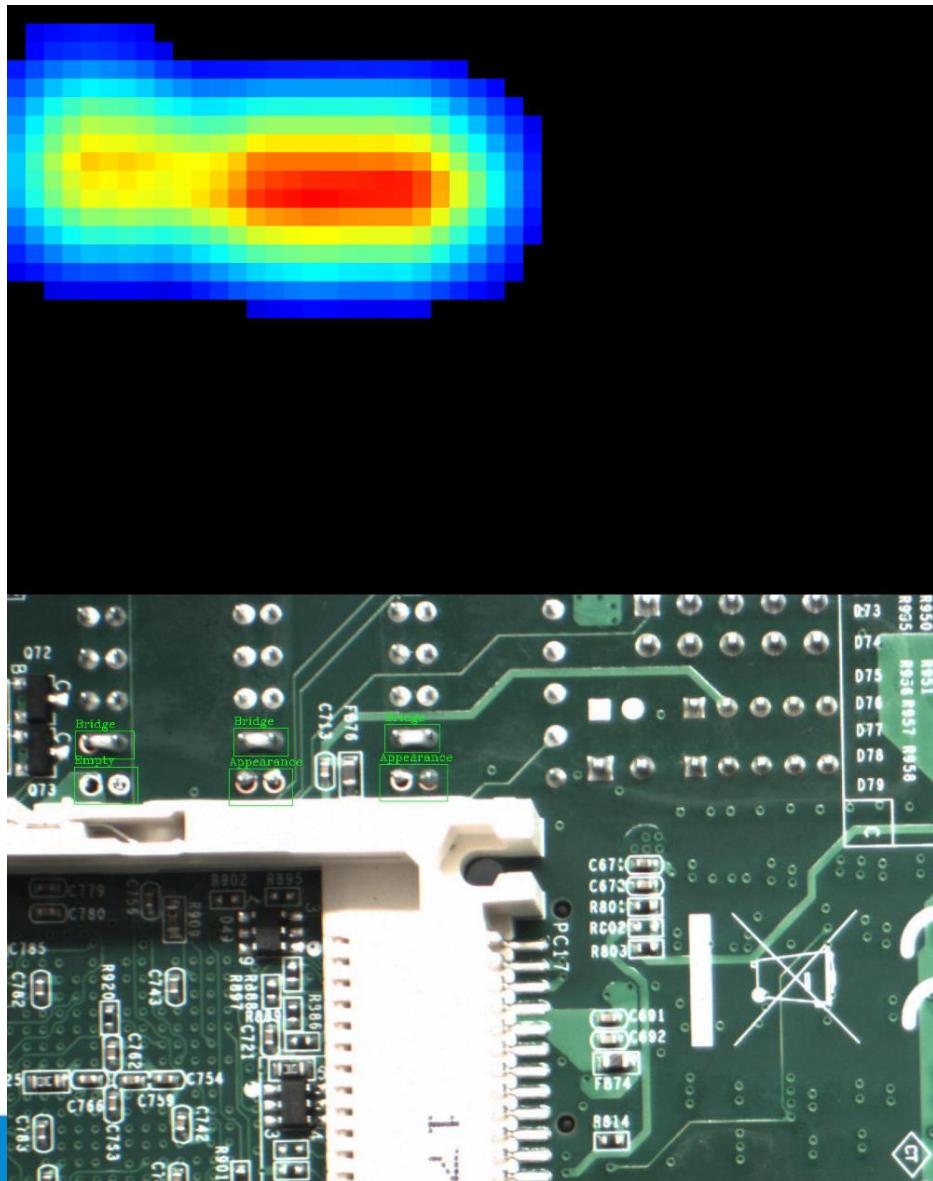
電路板上的缺陷偵測與分類

多種瑕疵同時存在，且人為錯誤在所難免



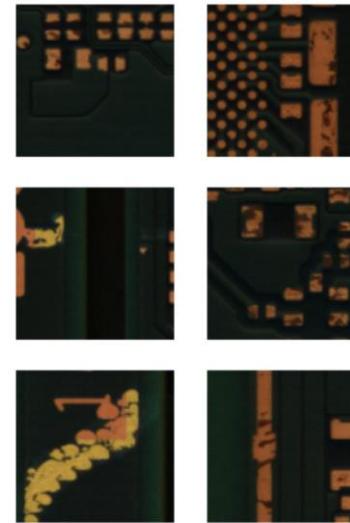
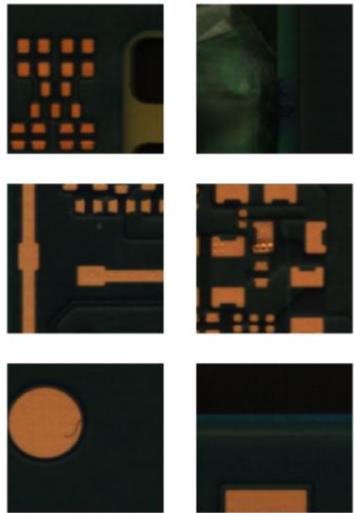
板子到底有沒有問題 - 是否應該人為介入確認

Classification Activation Map (CAM) : Why our network make this decision



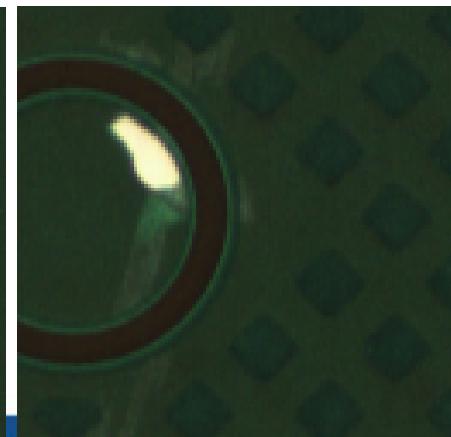
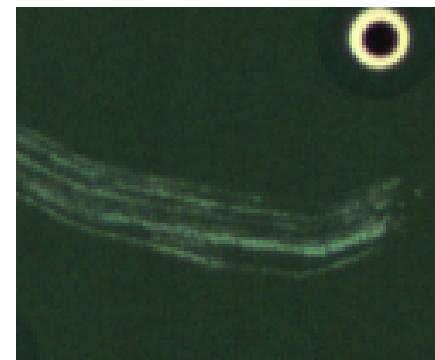
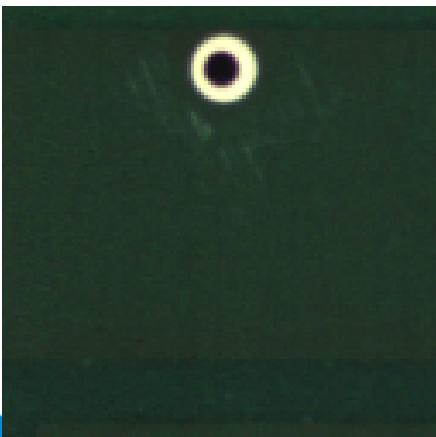
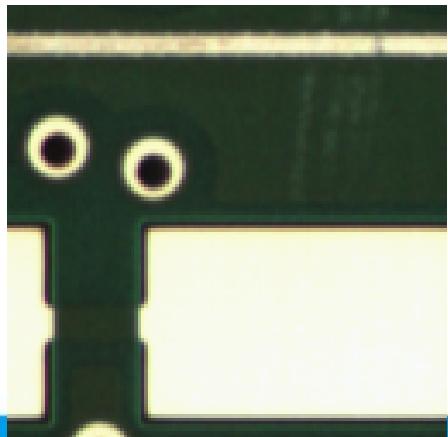
更多電路板的瑕疵檢測與分類問題

不同廠商 / 用途的板子 = 不同的缺陷與判定標準



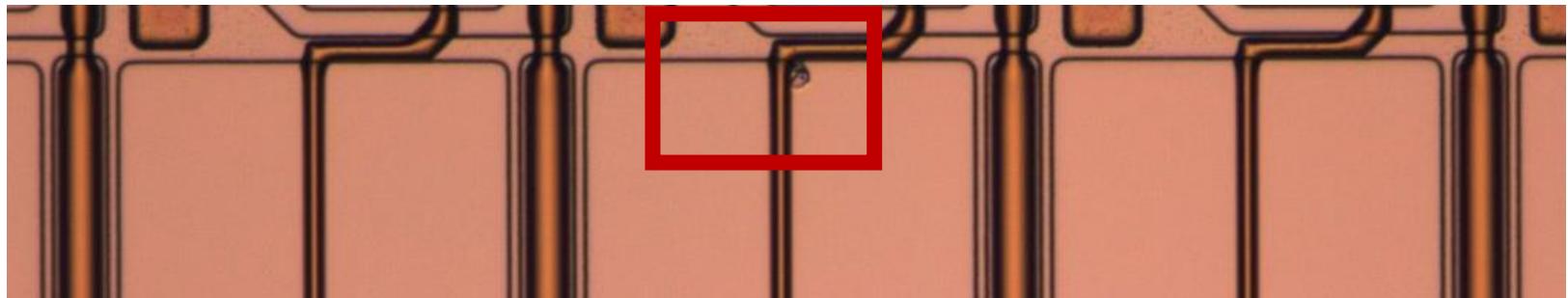
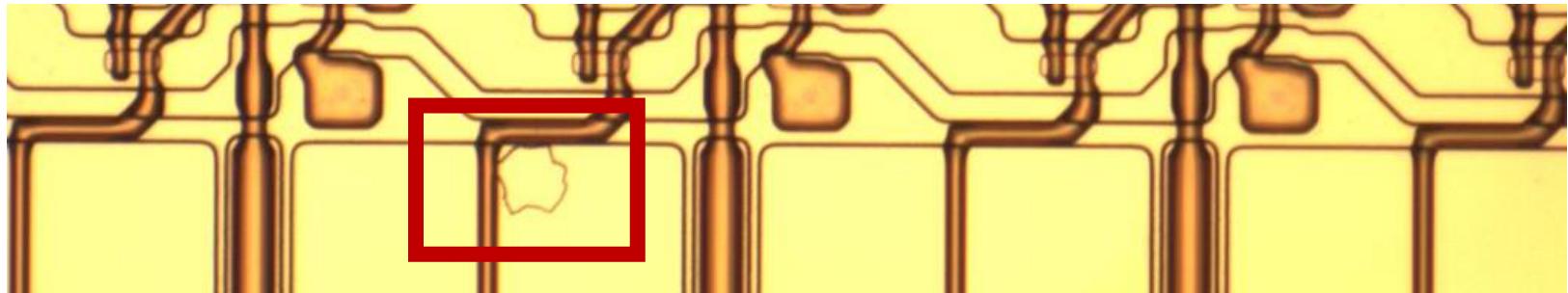
OK

NG



更多電路板的瑕疵檢測與分類問題

不同廠商 / 用途的板子 = 不同的缺陷與判定標準



視覺檢測：電子電路板缺陷偵測效益評估

傳統
人力
目檢

品質：根據複判初步統計，目檢人員漏網率至少為 **12.9%**

速度：目檢人員 8~10 位，每天約可檢查共約 3,000,000 張

深度
學習
系統

硬體設備：**中高階桌上型電腦 + NVIDIA GPU: 10 ~ 15 萬**

軟體：開源軟體 + 高度調校之深度學習模型

品質：模型漏網率控制在 1% 之下，目檢人員只需檢查原本總數之 10% 的圖片

速度：**8,640,000 張 / 天 = 100 張 / 秒**

Some interesting Cases

參數調控

配方選取與濃度調控

Materials combinations

- 找出最準確的打色配方，以解決公司因客人色樣要求所花時間與人力成本，應用其配方快速達到最相似的顏色以符合客戶的要求。



配方選取與濃度調控

What is the loss function?

$$\Delta E = \sqrt{\left(\frac{\Delta L}{lS_L}\right)^2 + \left(\frac{\Delta C}{cS_C}\right)^2 + \left(\frac{\Delta H}{S_H}\right)^2}$$

where

$$\Delta C = C_1 - C_2$$

$$C_1 = \sqrt{a_1^2 + b_1^2}$$

$$C_2 = \sqrt{a_2^2 + b_2^2}$$

$$\Delta H = \sqrt{\Delta a^2 + \Delta b^2 - \Delta C^2}$$

$$\Delta L = L_1 - L_2$$

$$\Delta a = a_1 - a_2$$

$$\Delta b = b_1 - b_2$$

$$S_L = \begin{cases} 0.511 & \text{if } L_1 < 16 \\ \frac{0.040975L_1}{1+0.01765L_1} & \text{if } L_1 \geq 16 \end{cases}$$

$$S_C = \frac{0.0638C_1}{1 + 0.0131C_1} + 0.638$$

$$S_H = S_C(FT + 1 - F)$$

$$T = \begin{cases} 0.56 + |0.2 \cos(H_1 + 168^\circ)| & \text{if } 164^\circ \leq H_1 \leq 345^\circ \\ 0.36 + |0.4 \cos(H_1 + 35^\circ)| & \text{otherwise} \end{cases}$$

$$F = \sqrt{\frac{C_1^4}{C_1^4 + 1900}}$$

$$H = \arctan\left(\frac{b_1}{a_1}\right)$$

$$H_1 = \begin{cases} H & \text{if } H \geq 0 \\ H + 360^\circ & \text{otherwise} \end{cases}$$

配方選取與濃度調控

What is the loss function?

$$\Delta E = \sqrt{\left(\frac{\Delta L}{lS_L}\right)^2 + \left(\frac{\Delta C}{\Delta C}\right)^2 + \left(\frac{\Delta H}{\Delta H}\right)^2}$$

where

$$\Delta C = C_1 - C_2$$

$$C_1 = \sqrt{a_1^2 + b_1^2}$$

$$C_2 = \sqrt{a_2^2 + b_2^2}$$

$$\Delta H = \sqrt{\Delta a^2 + \Delta b^2}$$

$$\Delta L = L_1 - L_2$$

$$\Delta a = a_1 - a_2$$

$$\Delta b = b_1 - b_2$$

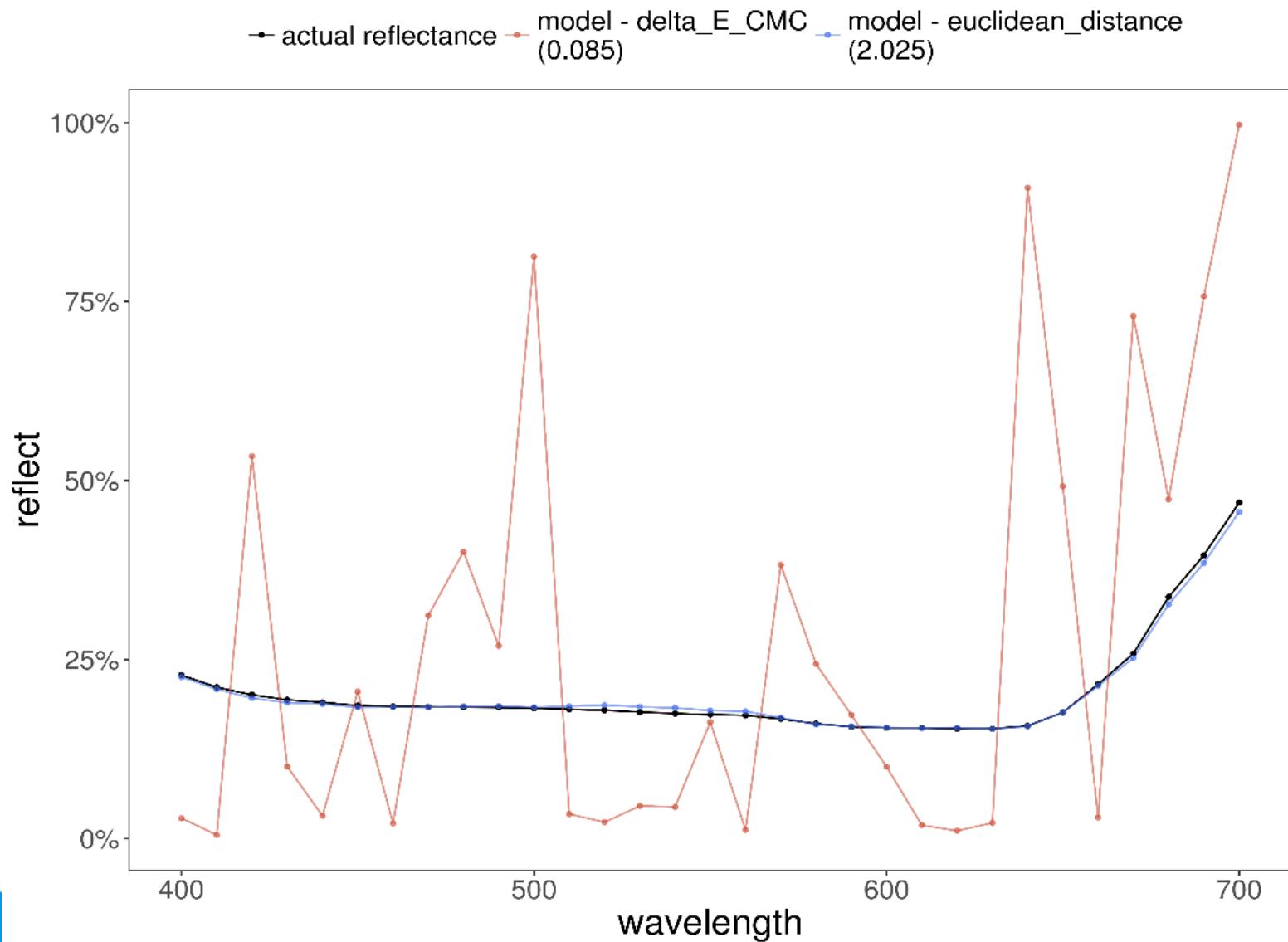


if $164^\circ \leq H_1 \leq 345^\circ$
otherwise

$$H_1 = \begin{cases} H & \text{if } H \geq 0 \\ H + 360^\circ & \text{otherwise} \end{cases}$$

原料配色 – 神秘的那道光?

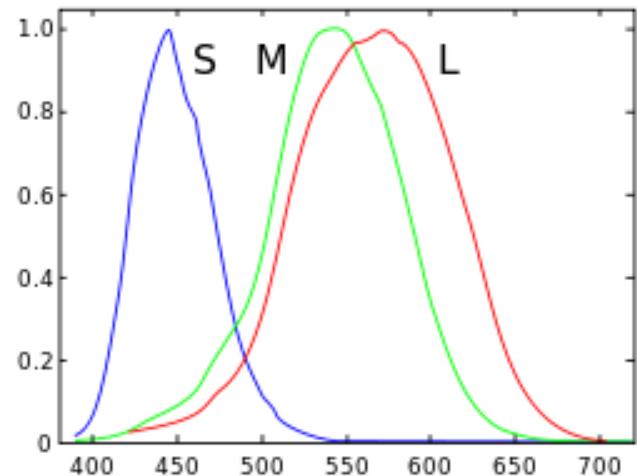
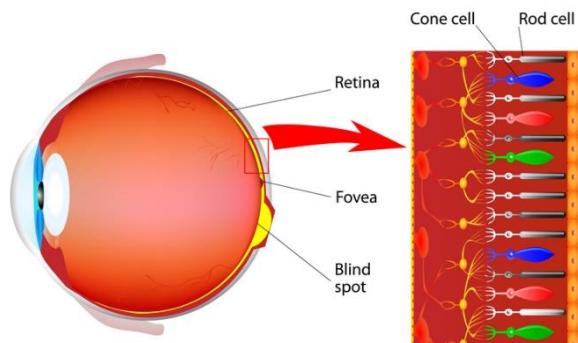
直觀的 Loss function = 詭異的 fitting result? : dECMC vs. eud.loss



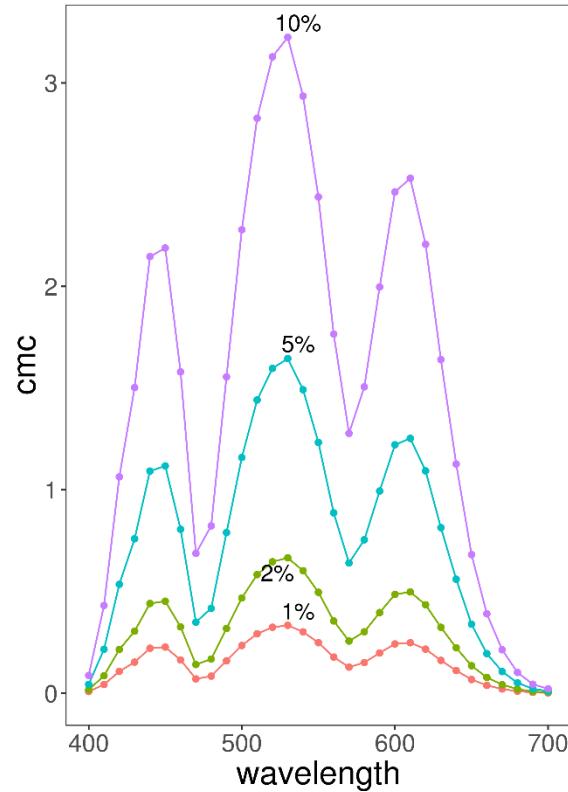
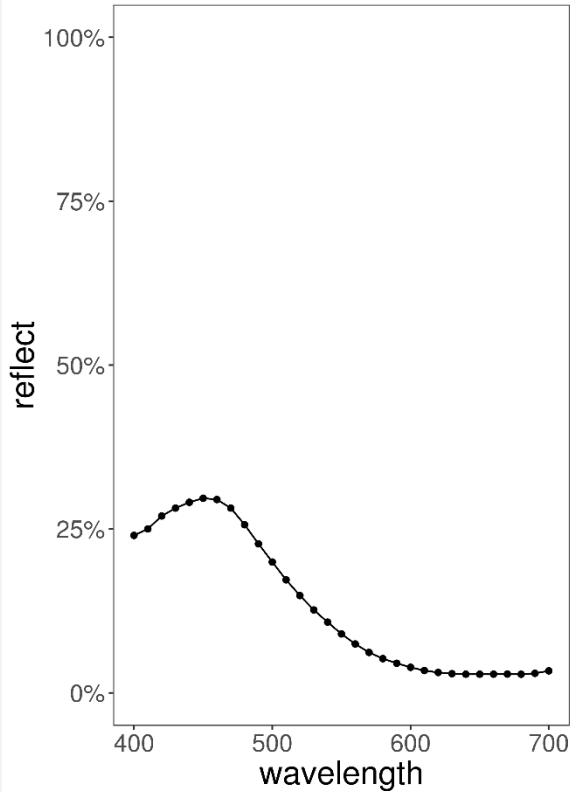
原料配色 – 神秘的那道光?

為何出現詭異的結果?

Photoreceptor cell

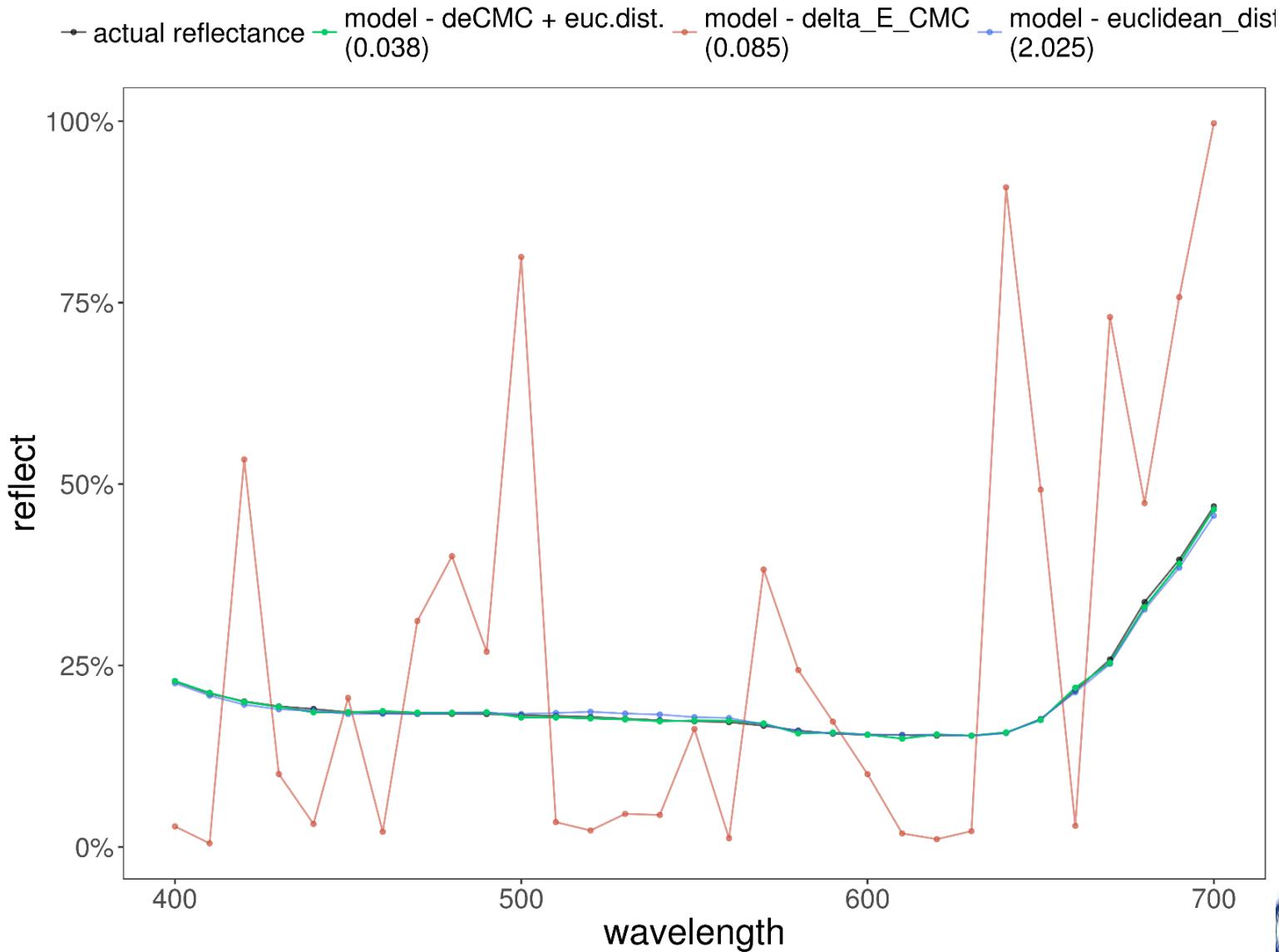


distribution of cmc
by adjust reflectance in each wavelength



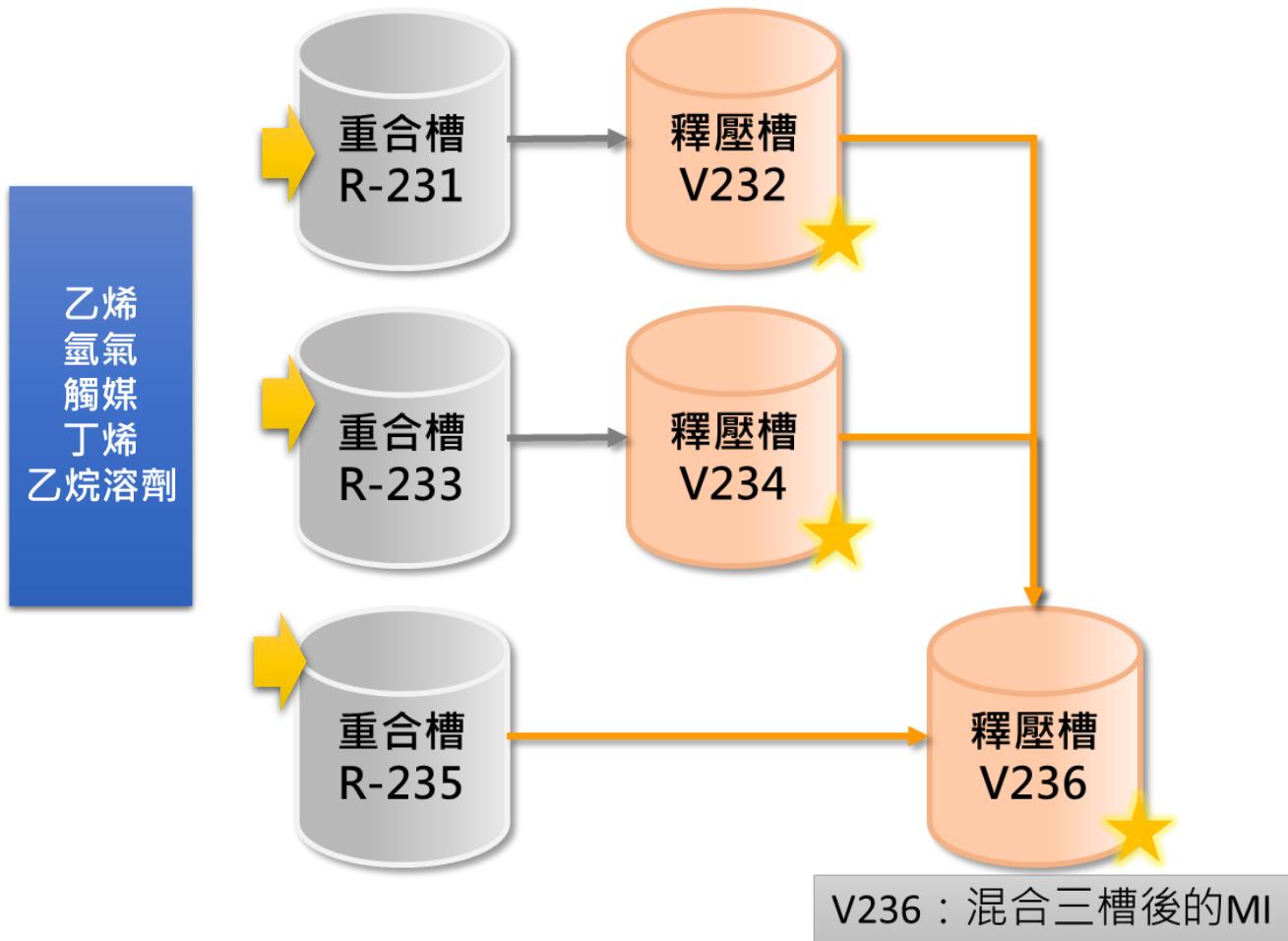
原料配色 – 神秘的那道光?

改成兩階段的 model fitting



化學製程產品的品質調控

Parameters optimization



化學製程產品的品質調控

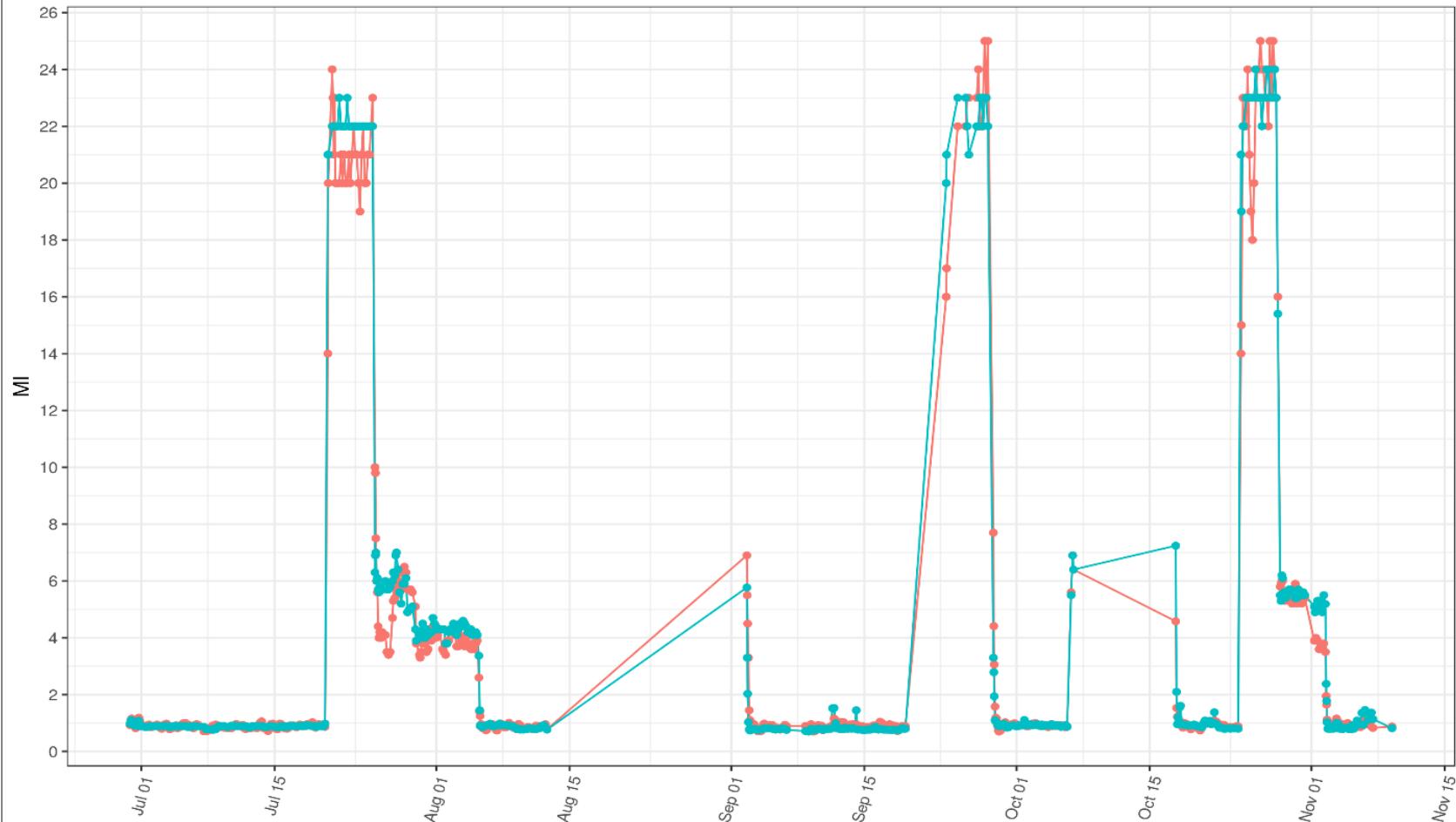
預測階段

MI 預測結果

時段: 2016-06-29 19:30:00 ~ 2016-11-09 11:00:00

● 實際值 ● 預測值

預測與實際值之相關係數可達
0.94 (相較過去傳統方法為 0.78)



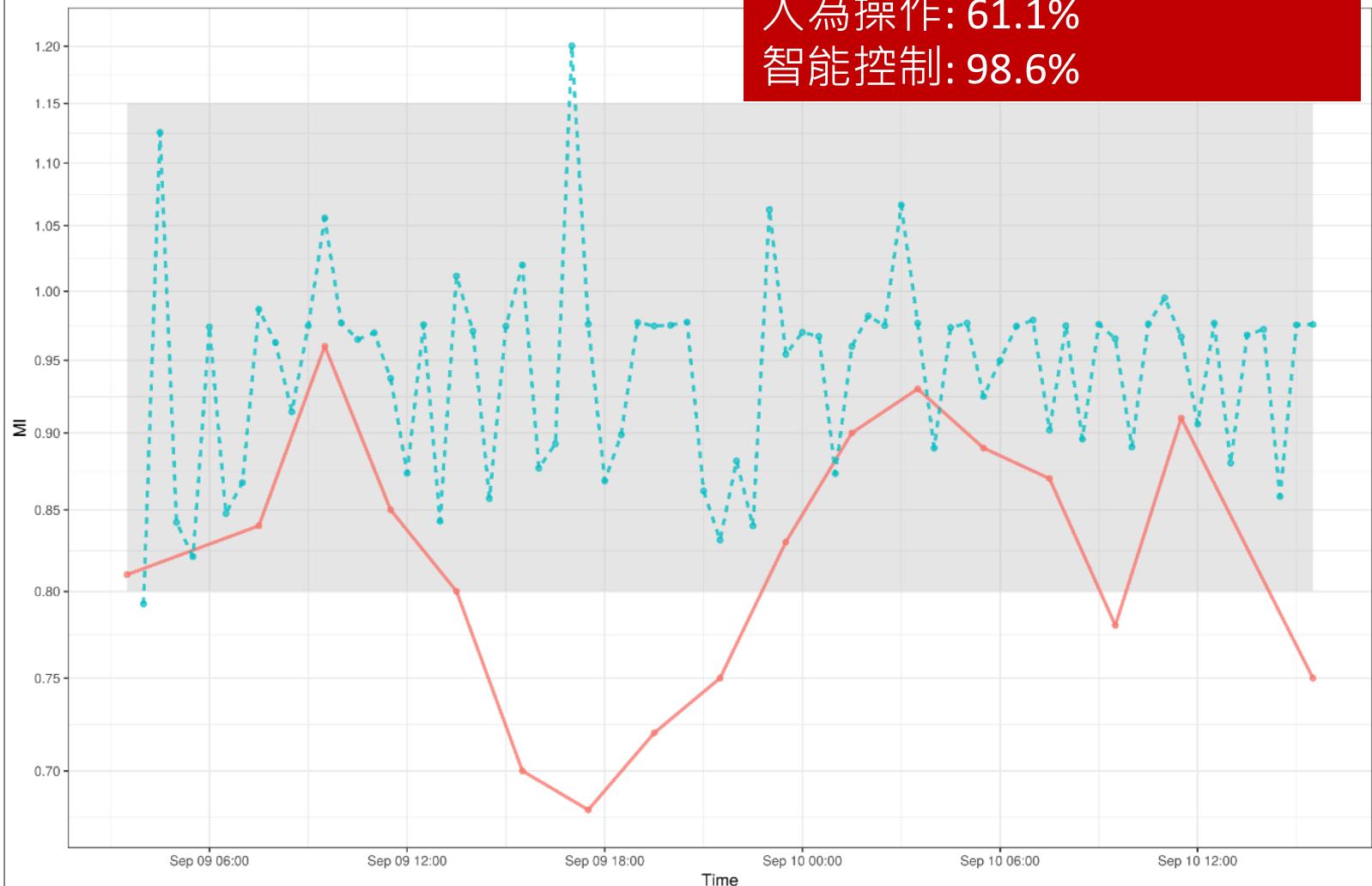
化學製程產品的品質調控

模擬控制階段

轉品階段 MI V234 之變化

轉品目標: 8010
時段: 2016-09-09 03:30:00 ~ 2016-09-10 15:30:00
人為操作之檢測結果 智能控制之預測結果

MI 在目標範圍時間比例
人為操作: 61.1%
智能控制: 98.6%



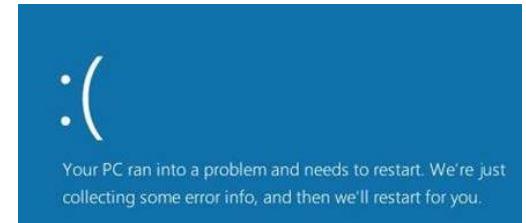
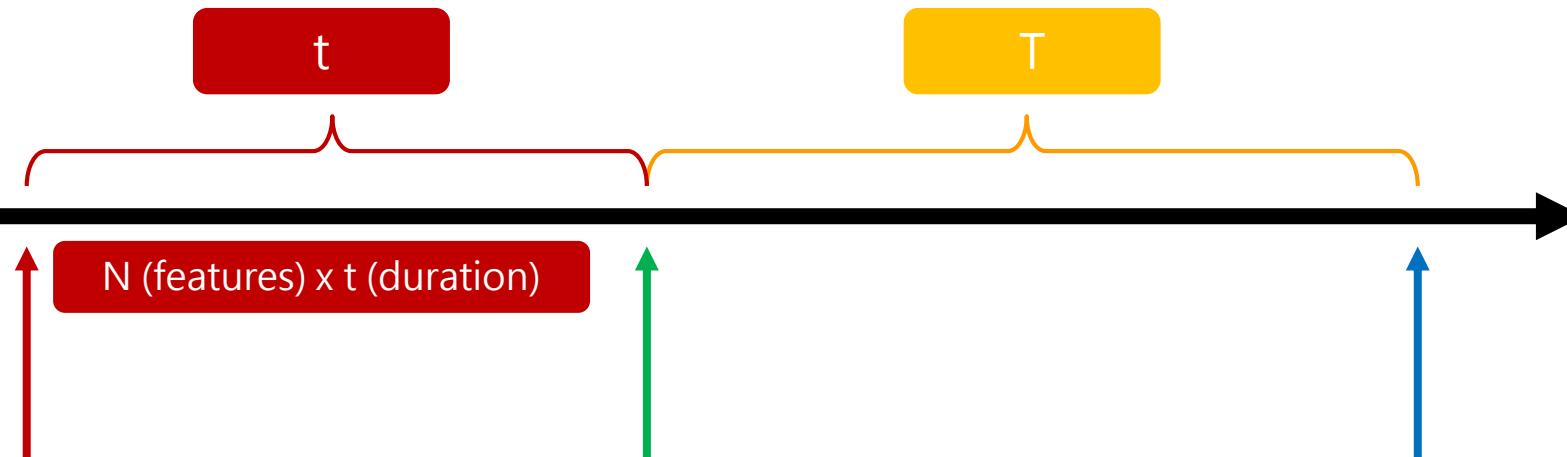
Some interesting Cases

預測性維護

提前預警 / 判斷機器何時會壞掉

可以當成分類問題？

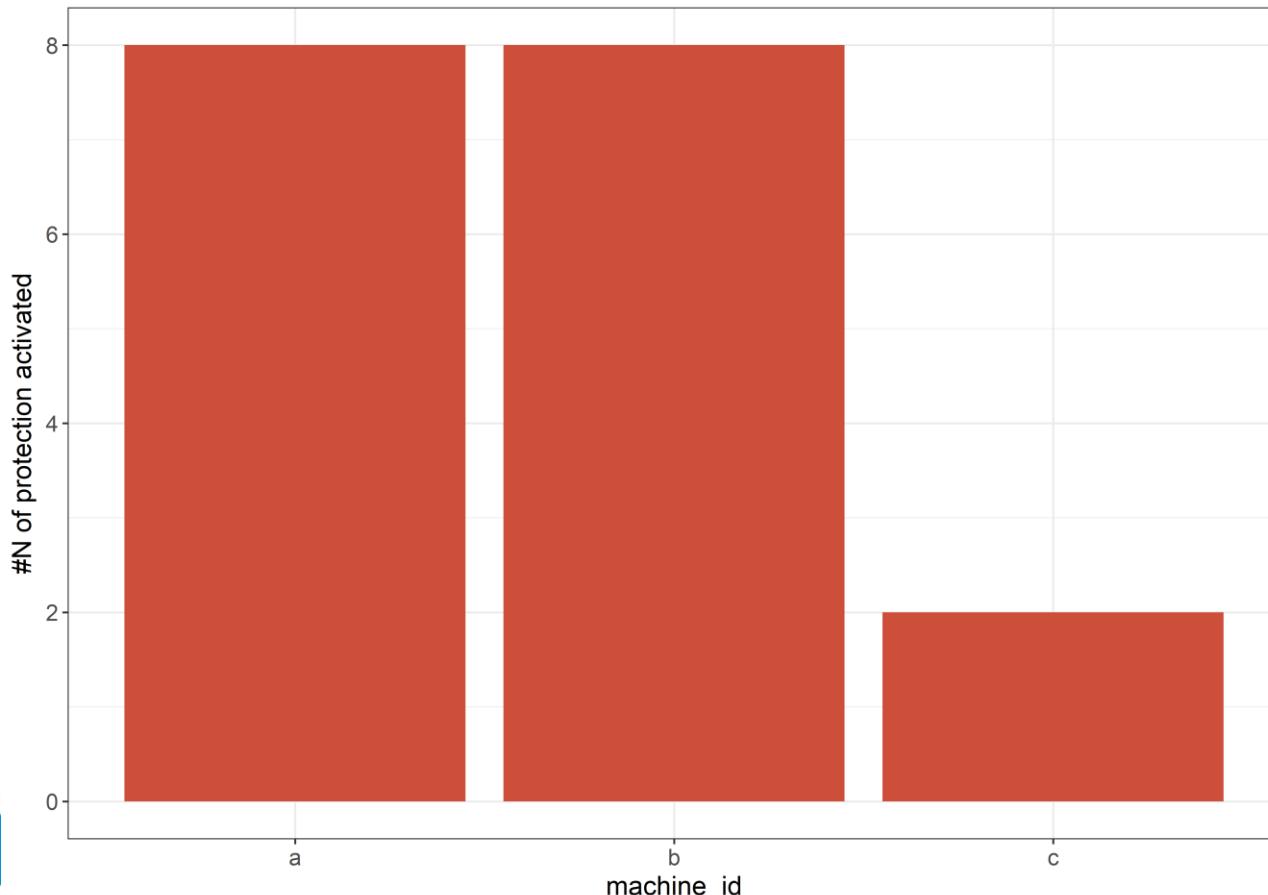
- 標準問題定義



提前預警 / 判斷機器何時會壞掉

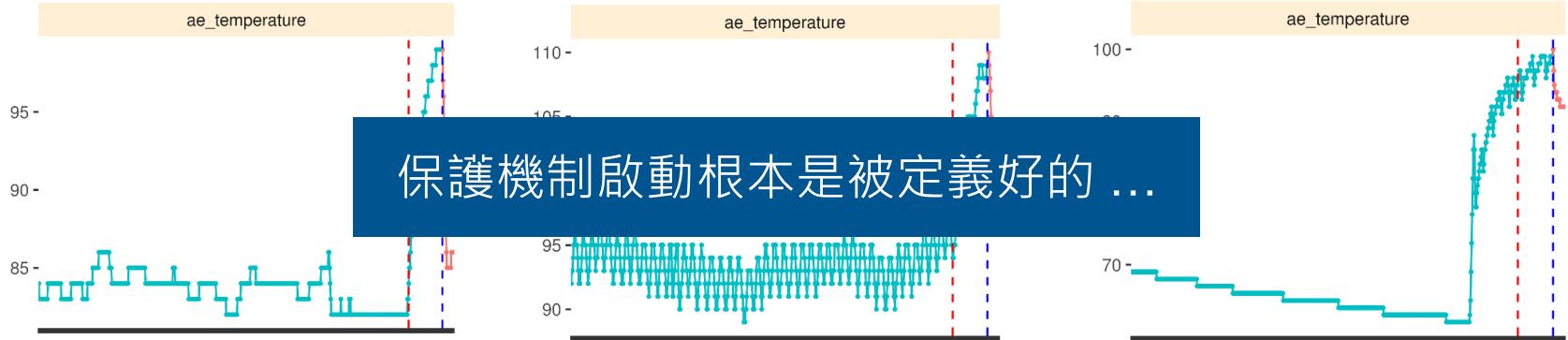
可以當成分類問題？

- 八個月來... 最多也只出現八次異常?!
 - 每 5 秒一筆資料, 八個月就是 ...五百多萬筆 ...
- “異常” 不是損壞, 只是保護機制 ...

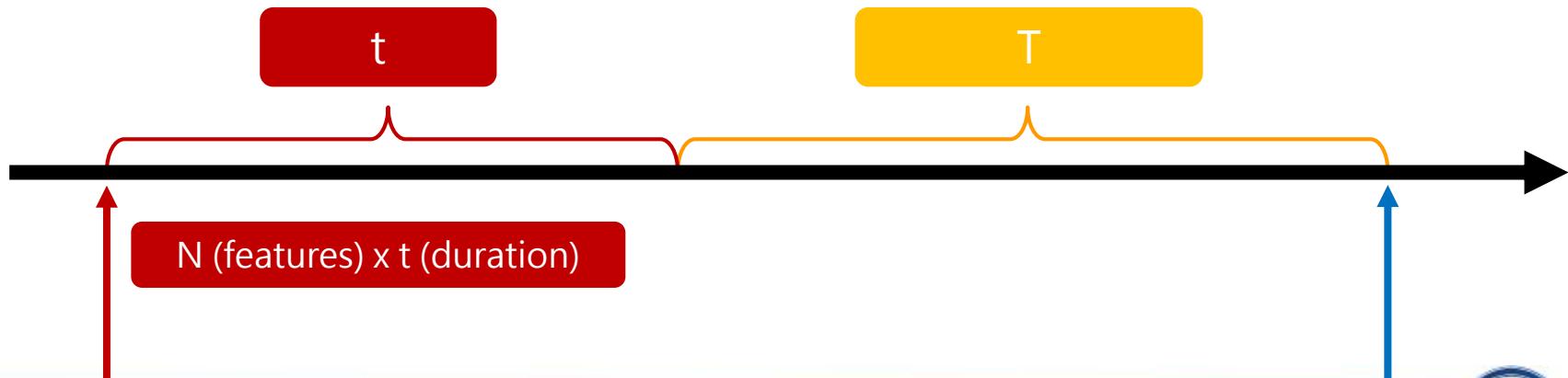


提前預警 / 判斷機器何時會壞掉

重新定義問題

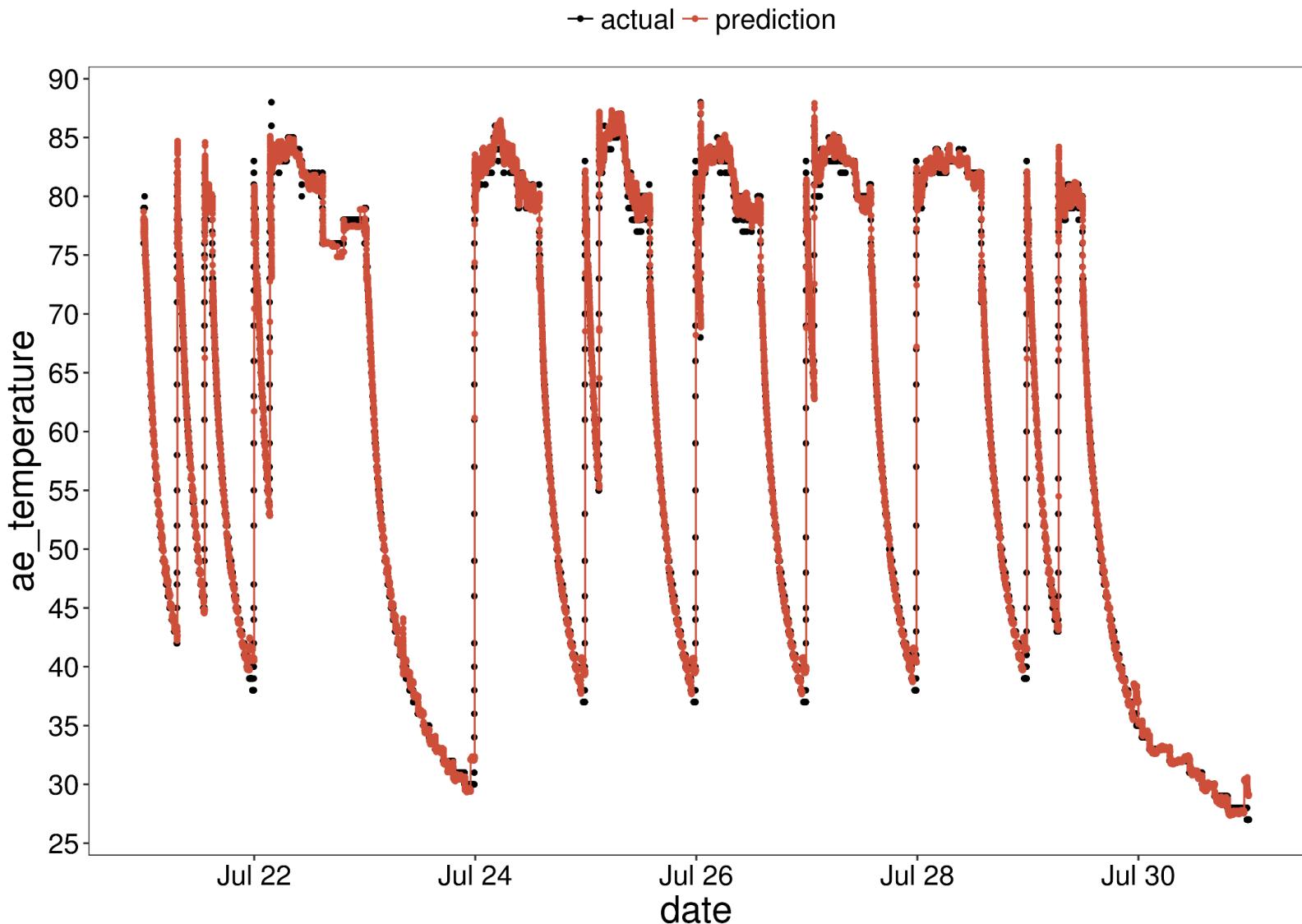


壓死駱駝的最後一根稻草什麼時候會出現？



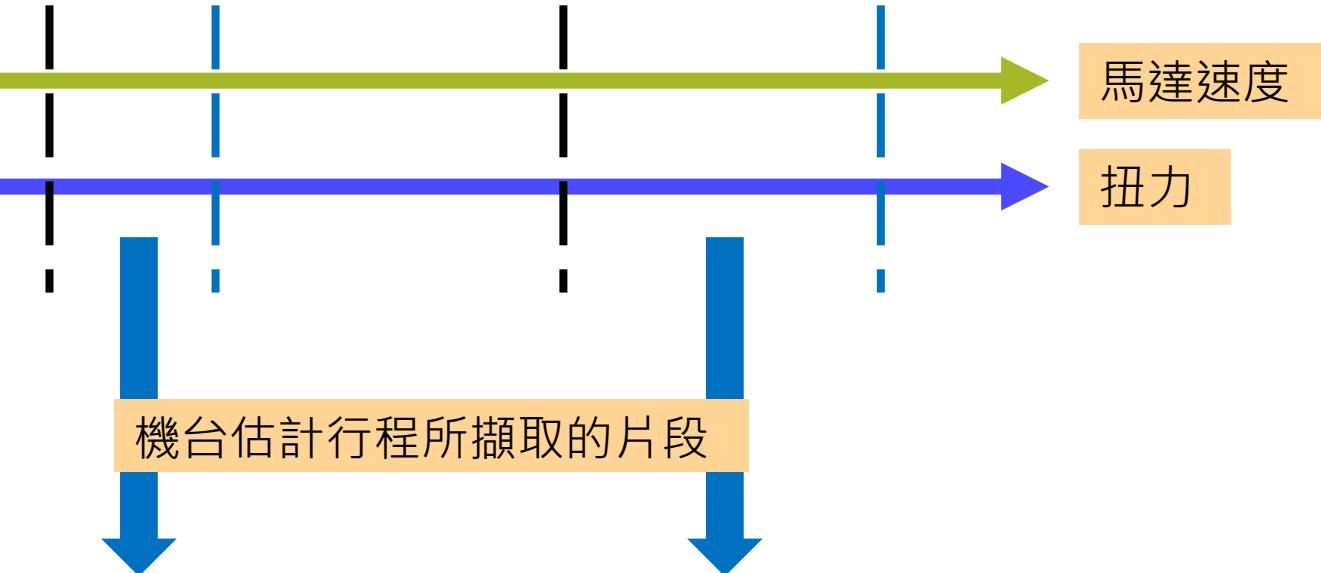
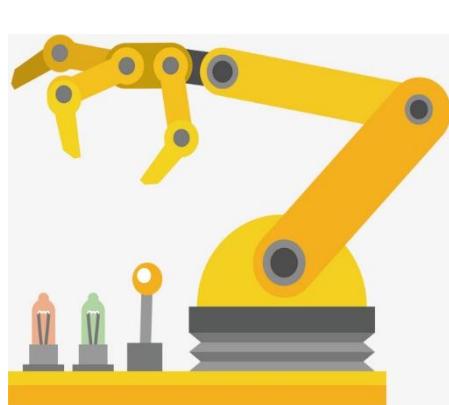
提前預警 / 判斷機器何時會壞掉

轉換目標：變成迴歸問題



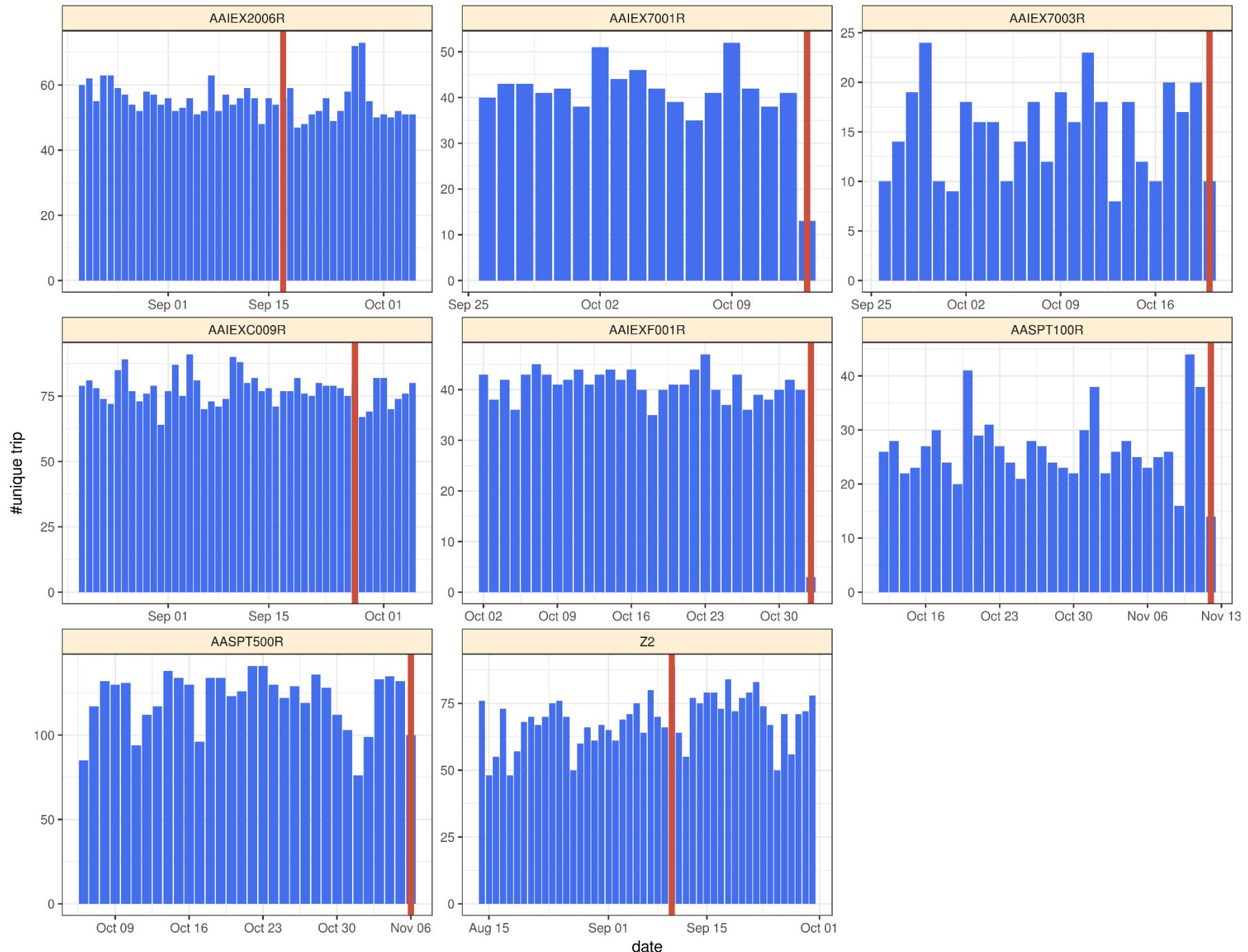
提前預警 / 判斷機器何時會壞掉

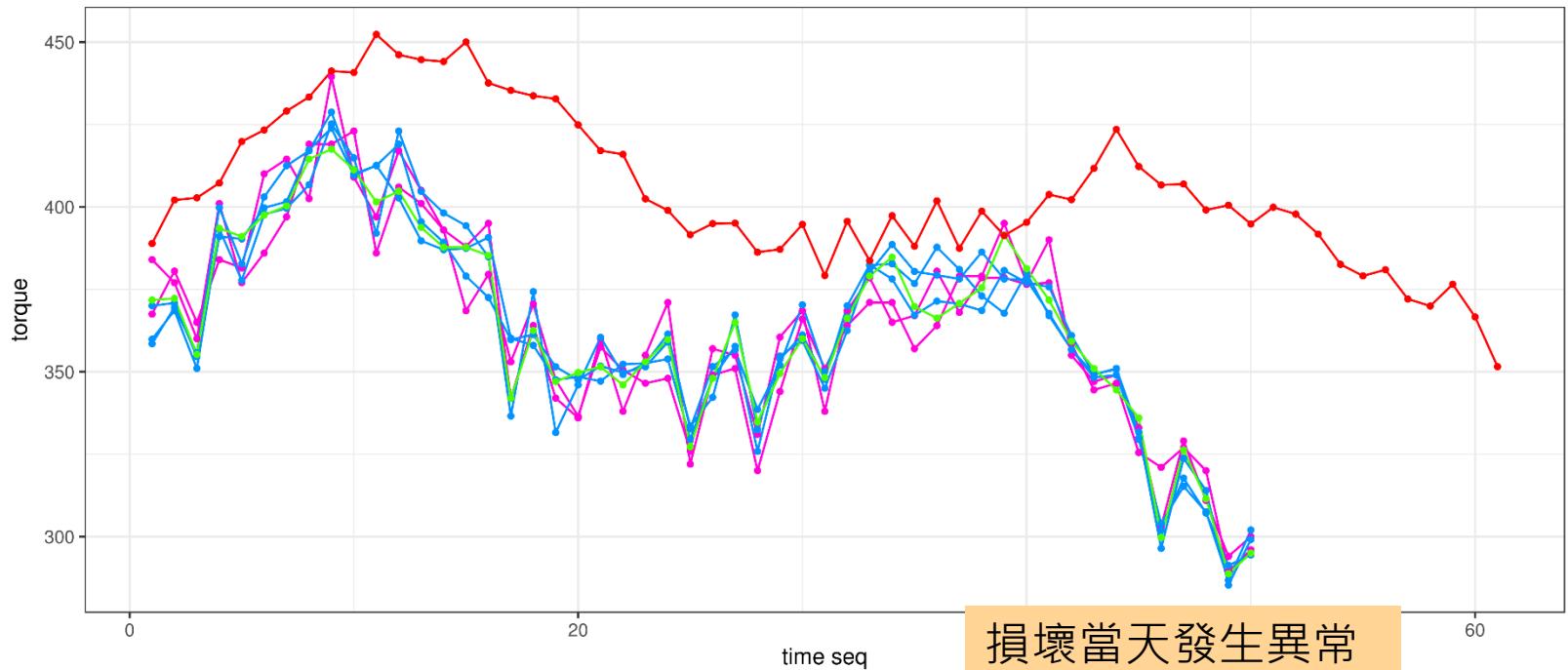
片段的連續性資料如何做起



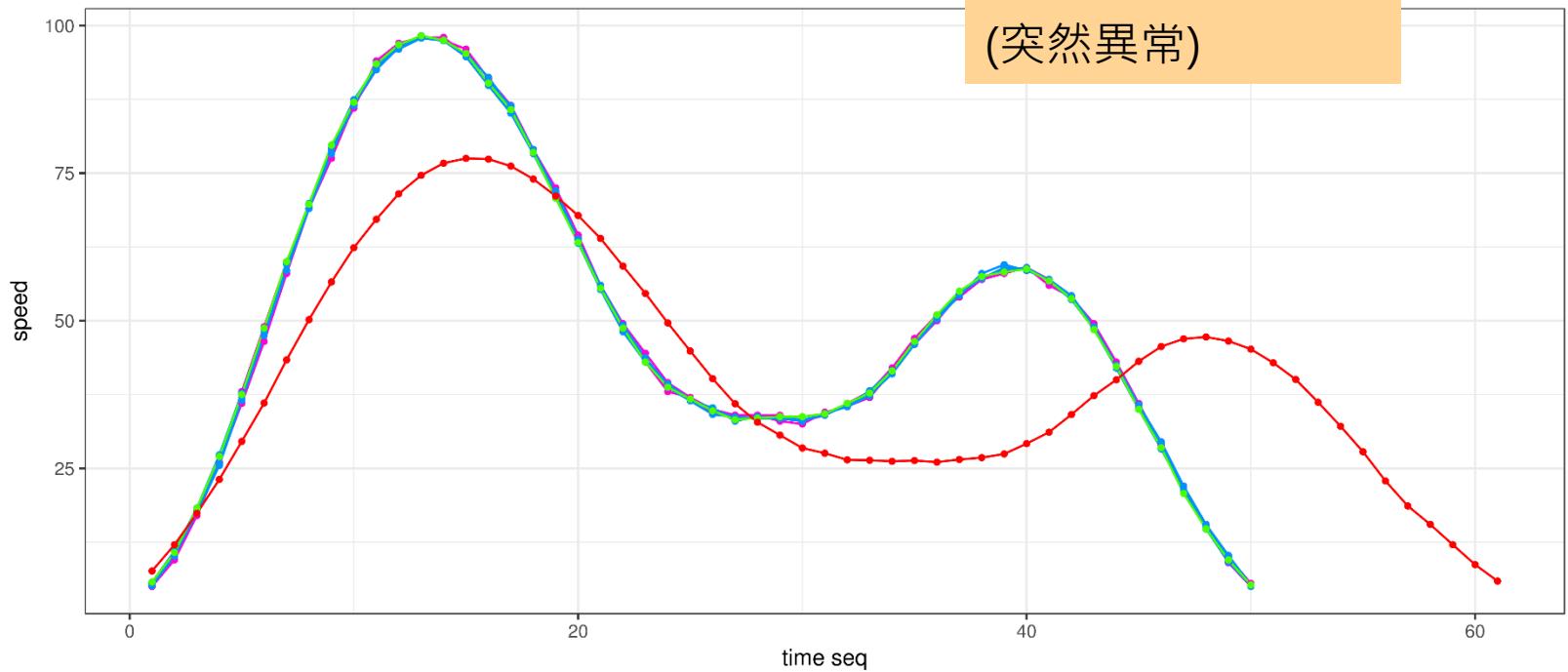
V1	V2	V3	V4	V5
2017-09-26 00:03:54.000	AXI_4_X	-42	351	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-36	308	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-31	296	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-26	300	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-21	257	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-16	180	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-12	150	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-10	167	CT200_225 MS-375_-400 U4
2017-09-26 00:03:54.000	AXI_4_X	-7	120	CT200_225 MS-375_-400 U4
2017-09-26 00:08:47.000	AXI_4_X	-5	-214	CT200_225 MS-375_-400 U4
2017-09-26 00:08:47.000	AXI_4_X	-8	-197	CT200_225 MS-375_-400 U4
2017-09-26 00:08:47.000	AXI_4_X	-10	-253	CT200_225 MS-375_-400 U4
2017-09-26 00:08:47.000	AXI_4_X	-13	-265	CT200_225 MS-375_-400 U4
2017-09-26 00:08:47.000	AXI_4_X	-16	-300	CT200_225 MS-375_-400 U4
2017-09-26 00:08:47.000	AXI_4_X	-20	-334	CT200_225 MS-375_-400 U4

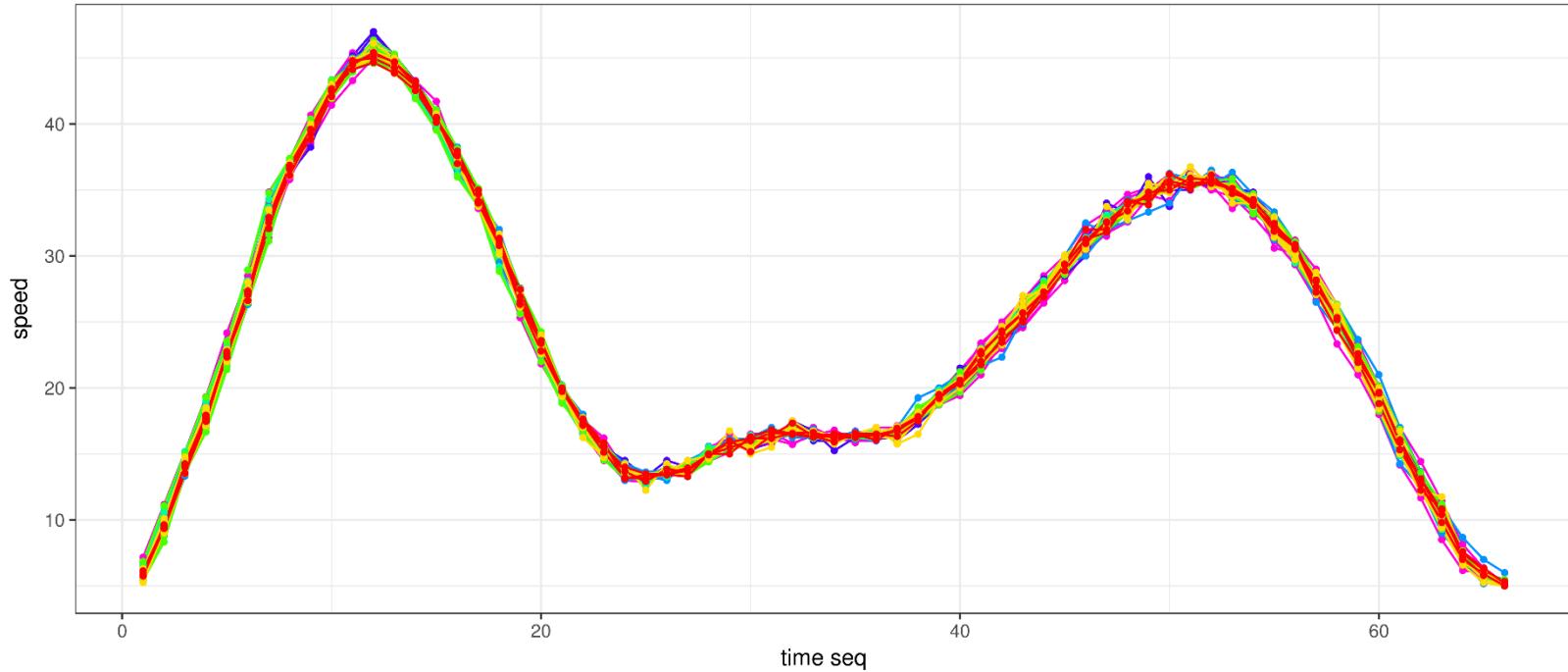
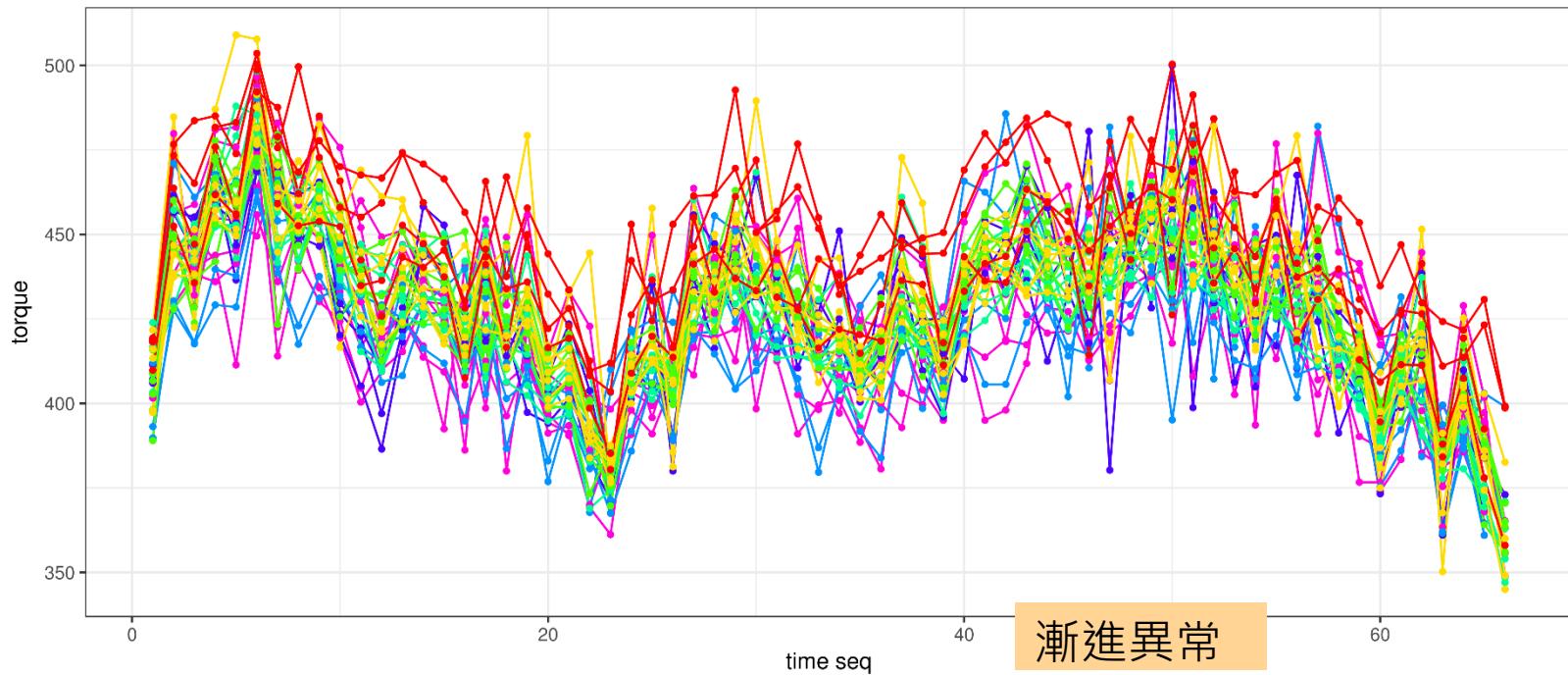
Numbers of unique trip per day



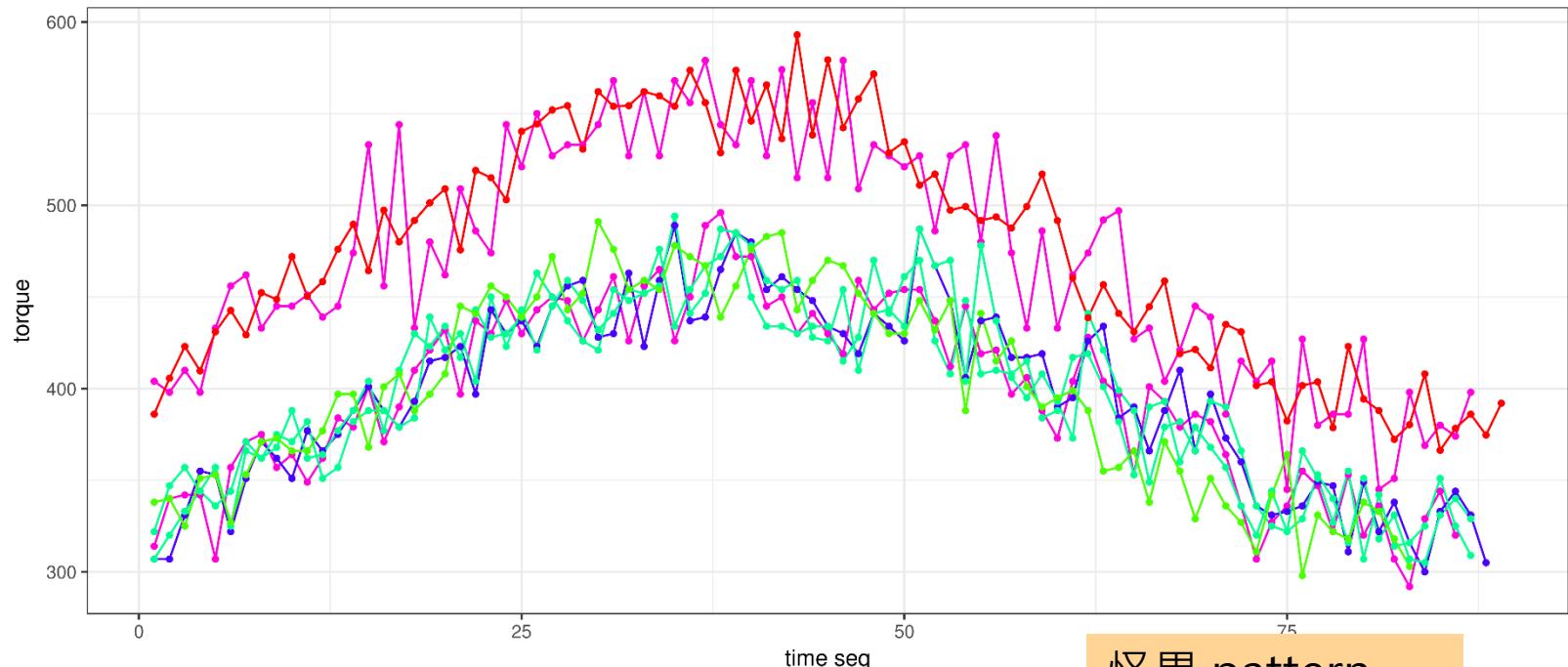


損壞當天發生異常
(突然異常)

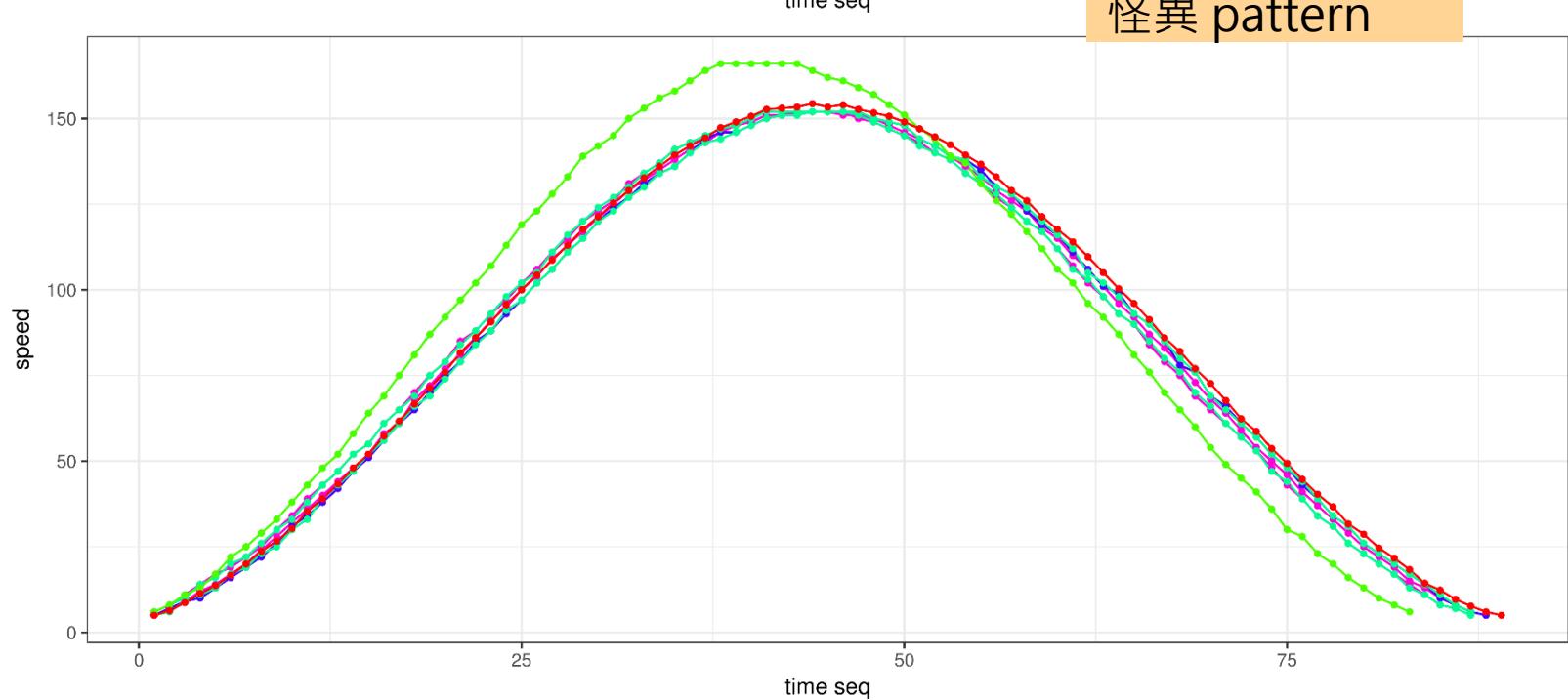




39



怪異 pattern

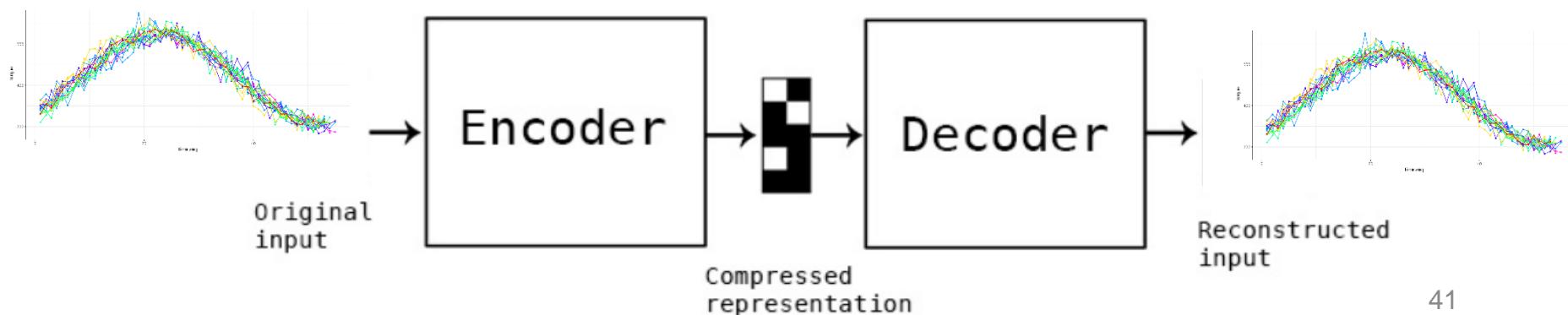
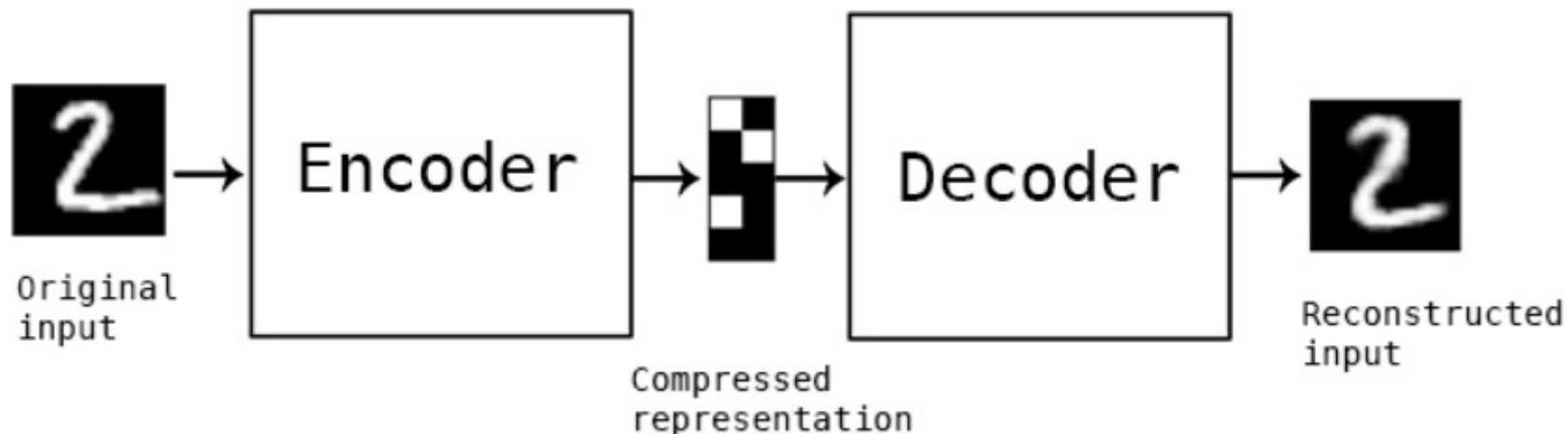


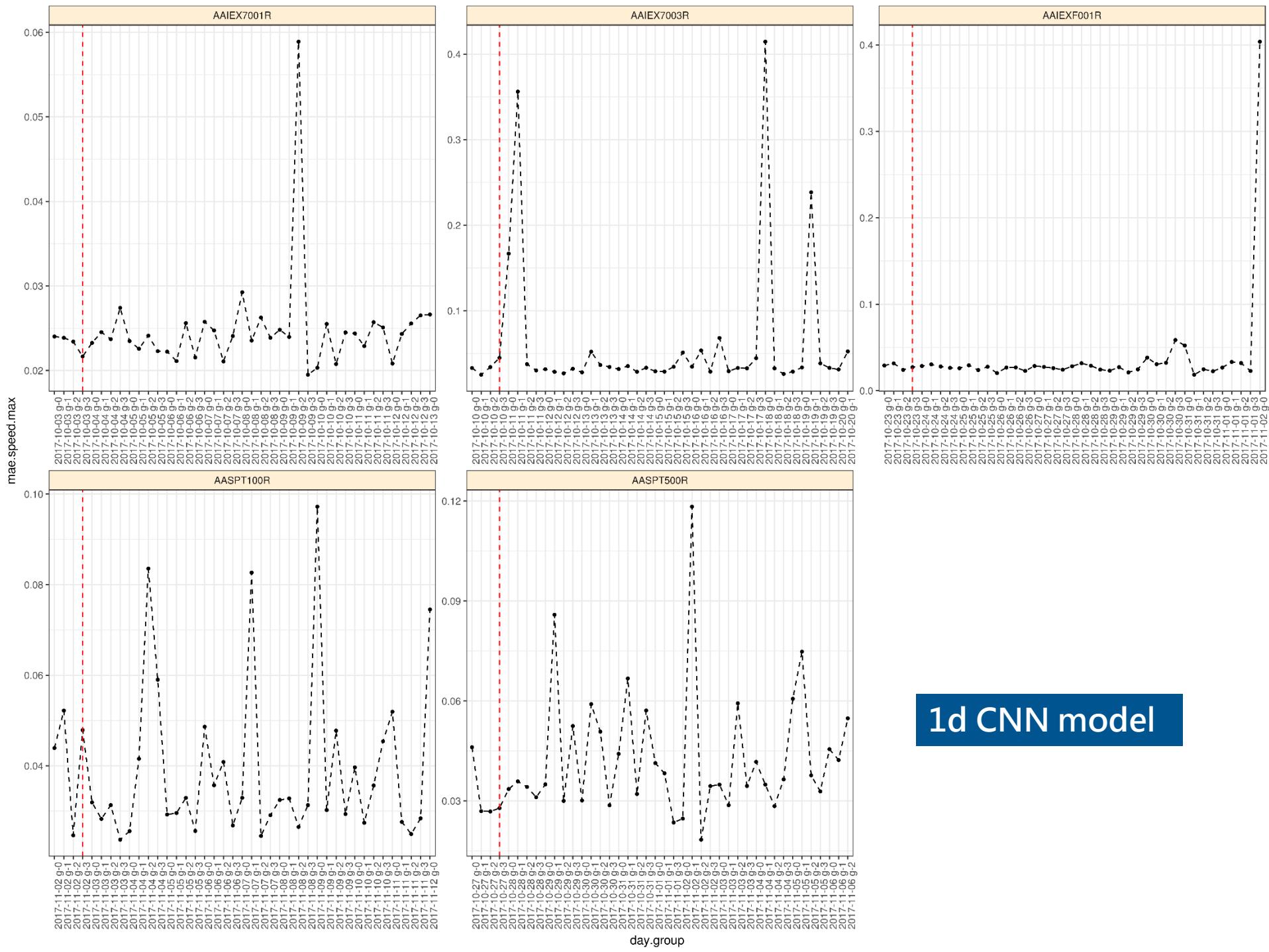
40

Anomaly detection with deep learning

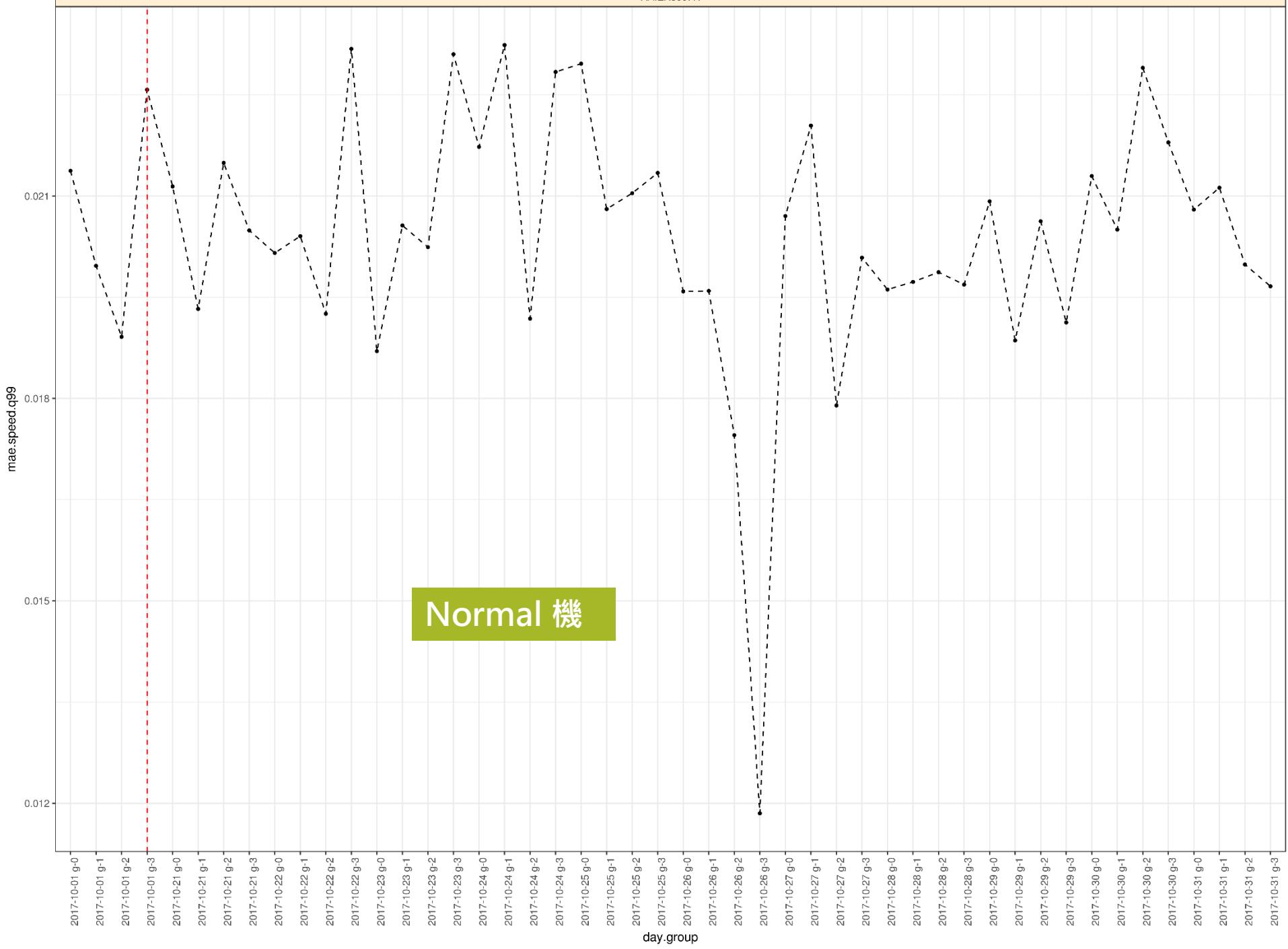
Sequences reconstruction by an auto-encoder

- 透過 deep neural network 將輸入經過編碼 (encoding) 後，用轉換後的 representation 重建出輸入值



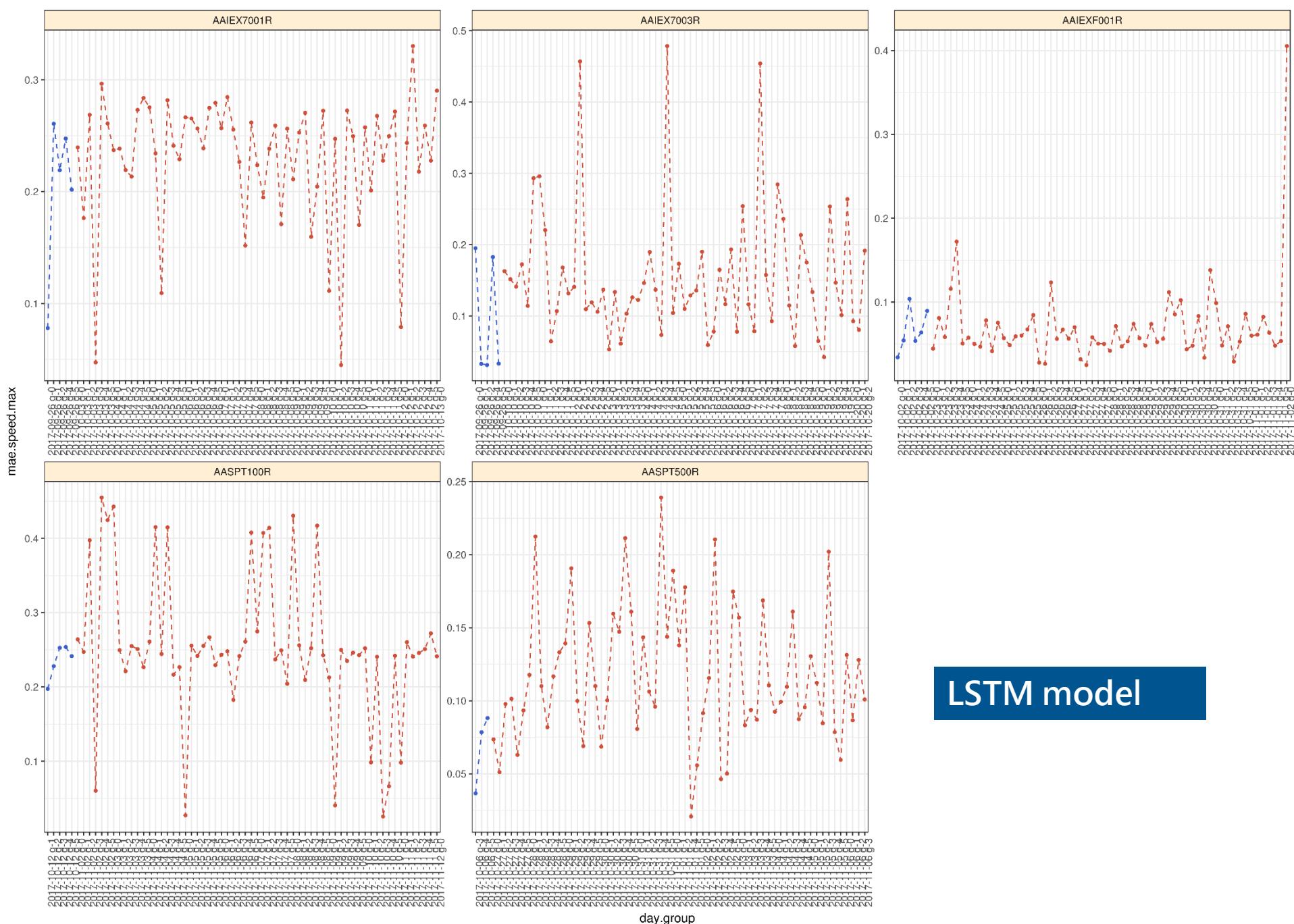


1d CNN model



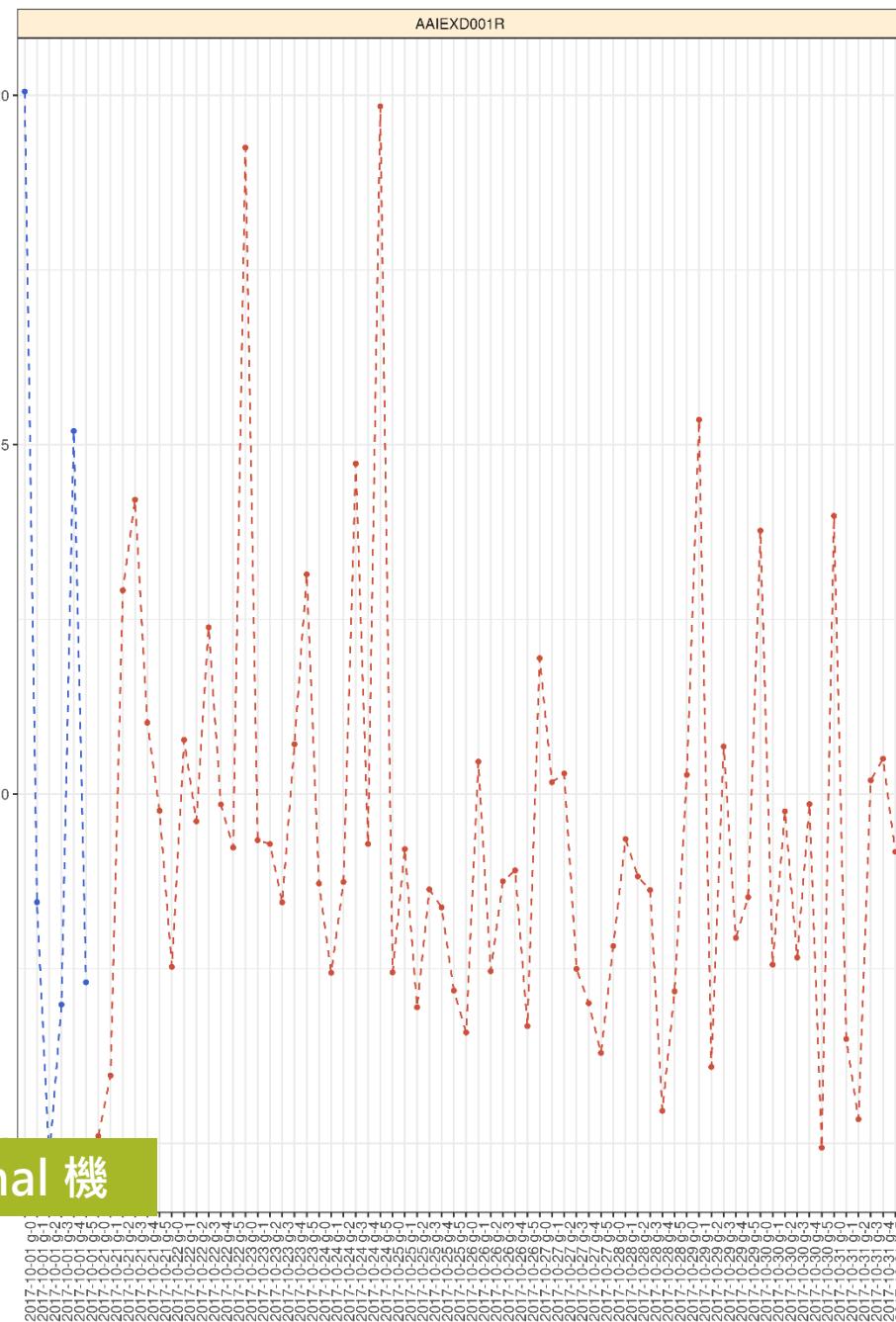
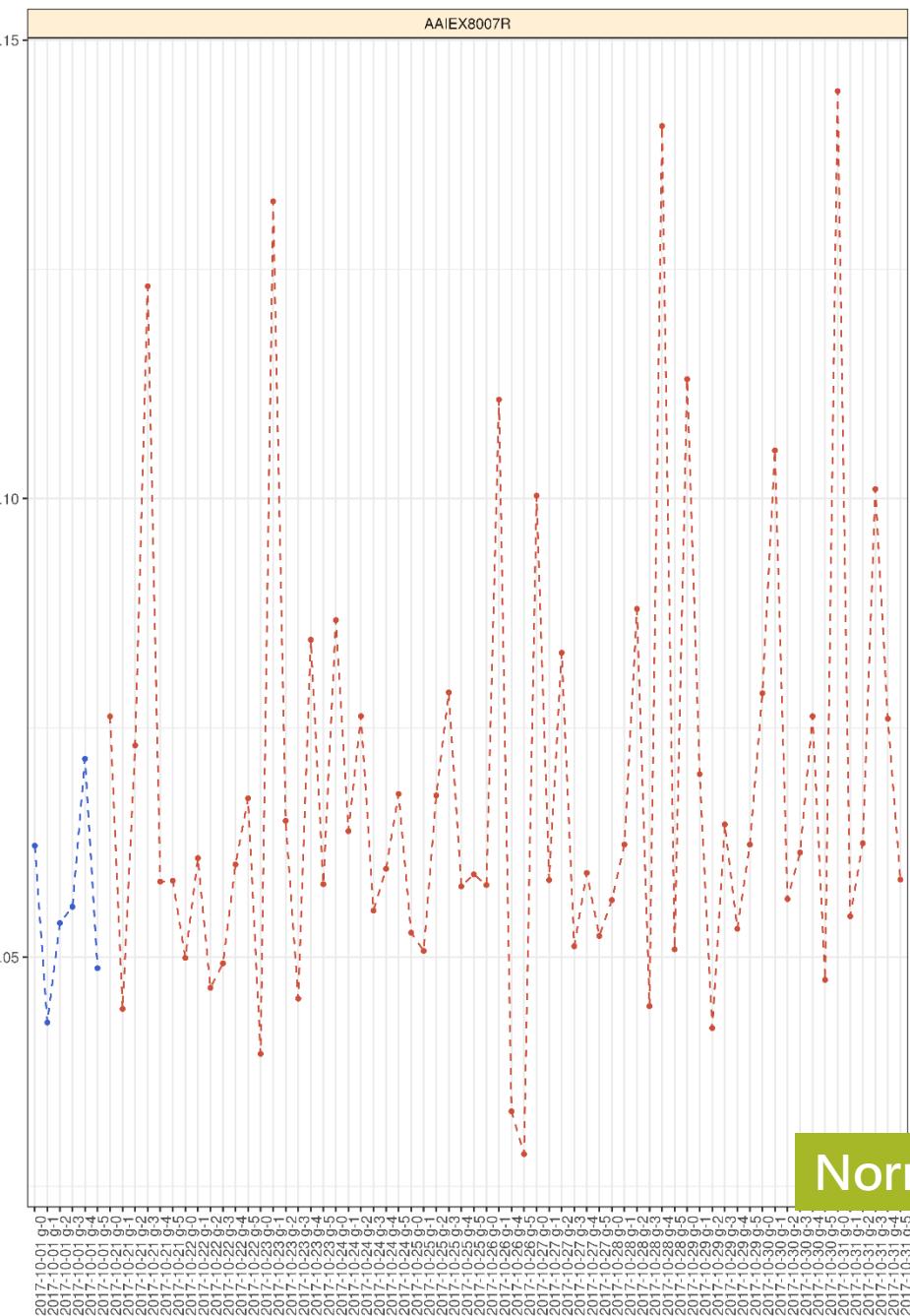
Normal 機

colour — normal — ruin



LSTM model

colour — blue first day — red last 10 days



Normal 機



拿 pre-trained model 或是參考網路上的範例很容易
但是 ...

問對問題、了解的每個模型限制、山不轉路轉
才能在不完美的現實世界中解決各式各樣奇怪問題

Thank You!