

台塑公司

②

油品雙烯值AI分析

檢驗中心 彙編

2020年9月26日

【密】 【會後收回】

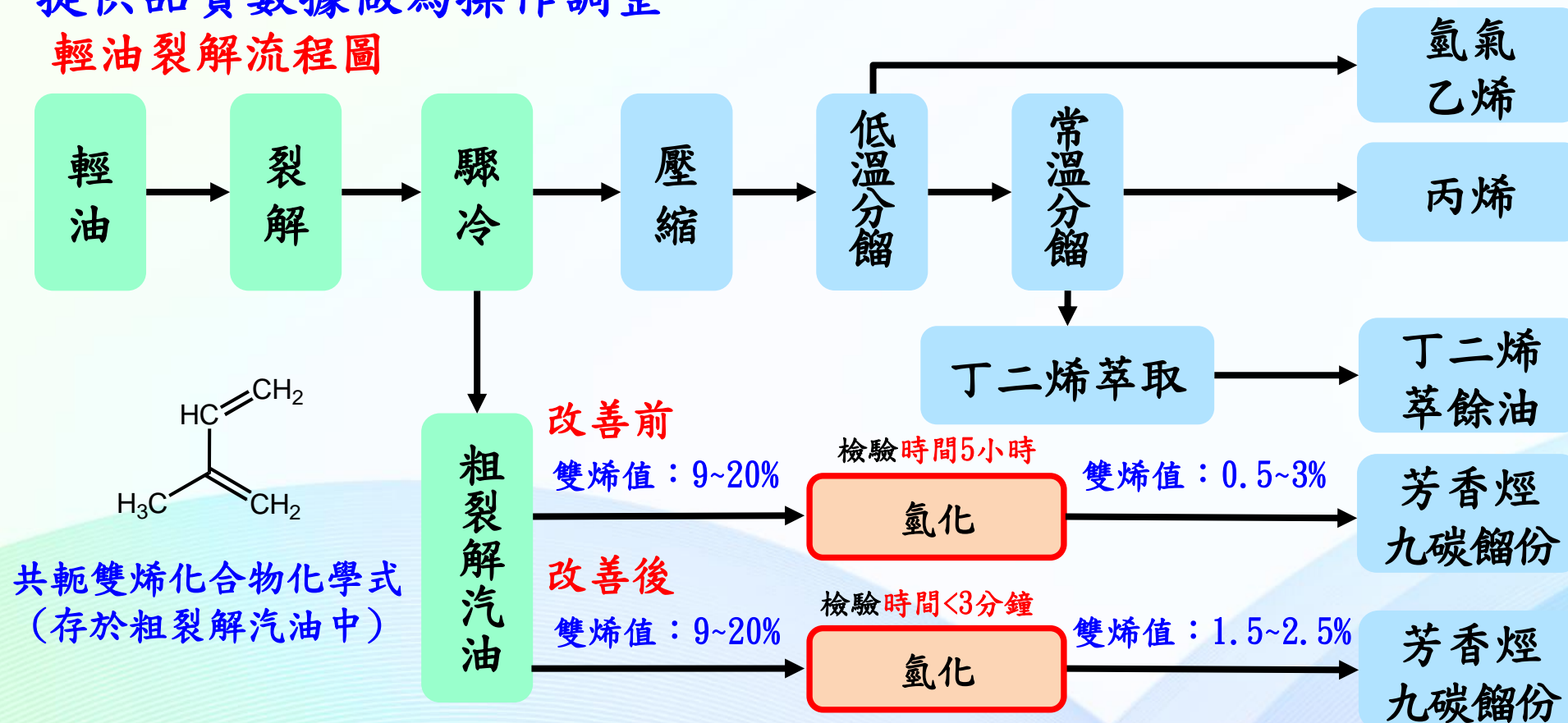
目錄

- 1. 改善動機
- 2. 檢驗流程
- 3. 解決方案
- 4. AI模型開發歷程
- 5. 各階段詳細說明
- 6. 效益說明
- 7. 後續推動事項

1. 改善動機

- 台塑檢驗中心負責台塑石化公司烯烴事業部輕油、粗裂解汽油、乙烯、丙烯、丁二烯、芳香烴等品質檢驗作業。
- 輕油經裂解、驟冷後所產生之粗裂解汽油，存在共軛雙烯化合物，檢驗時間約5小時，易使該類化合物聚合，影響後端產品之品質。
- 本案針對粗裂解汽油中雙烯值檢驗方法進行縮短檢驗時間改善，即時提供品質數據做為操作調整。

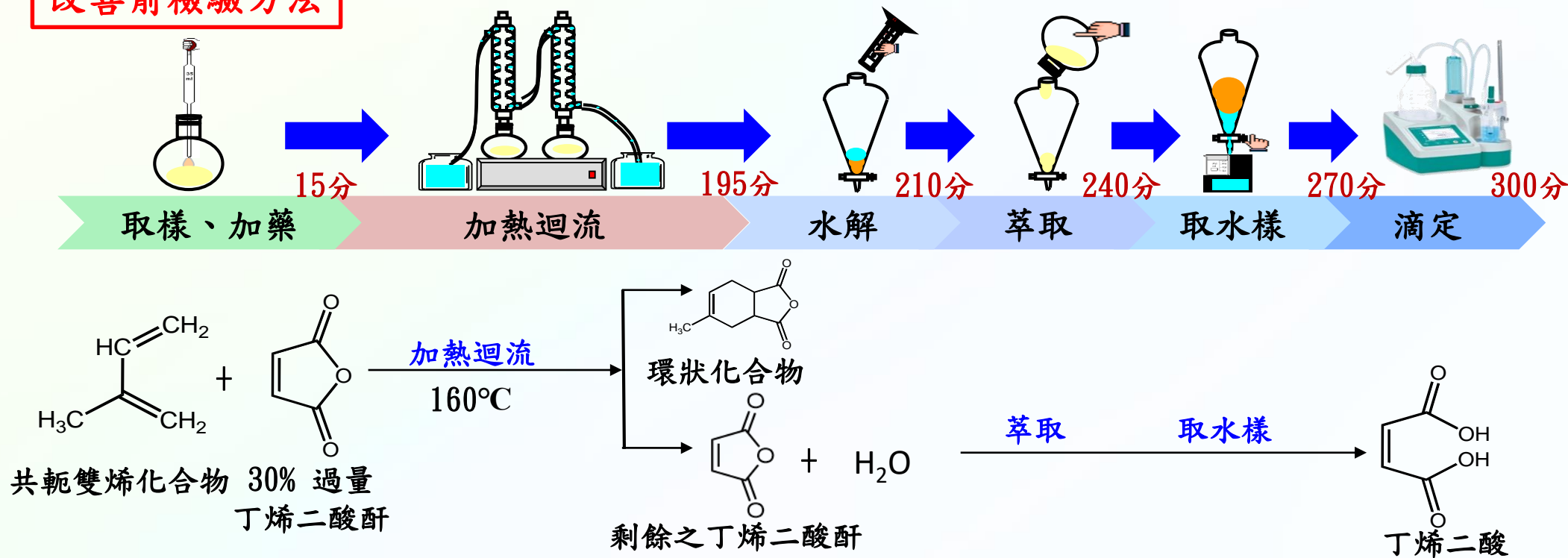
輕油裂解流程圖



2. 檢驗流程

➤ 雙烯值為分析汽油氫化區中共軛雙烯化合物含量，作為氫化操作的依據。

改善前檢驗方法

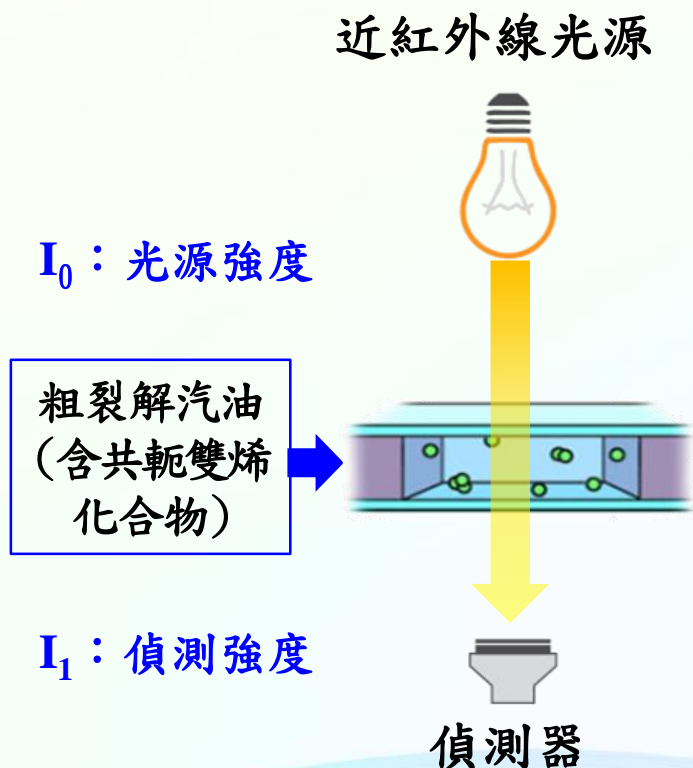


改善後檢驗方法



3.1 解決方案

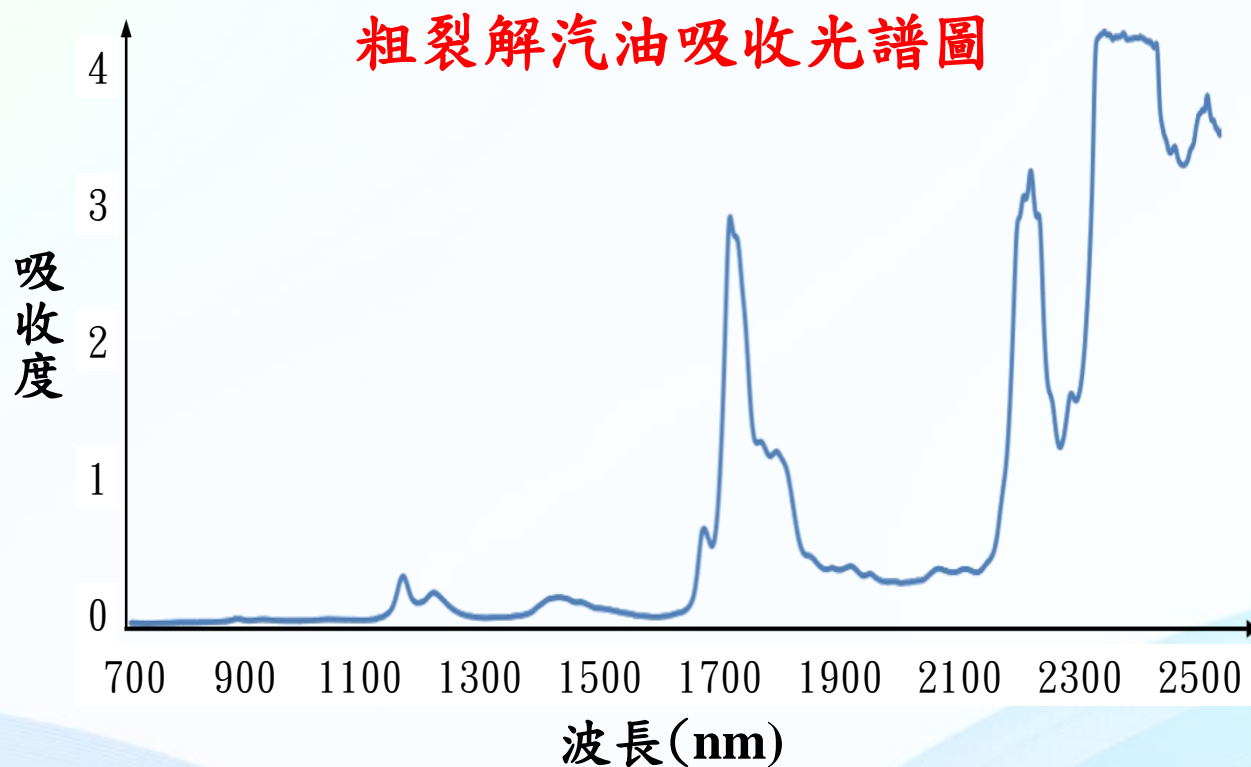
- 粗裂解汽油包含100種以上 $C_5 \sim C_{12}$ 碳氫化合物，利用碳氫化合物（含共軛雙烯化合物）吸收近紅外線能量特性，產生吸收光譜圖，再由AI判讀共軛雙烯化合物的濃度。



吸收度公式= $\log(I_0 / I_1)$

舉例： $I_0=100$ 燭光、 $I_1=1$ 燭光

吸收度= $\log(100 \text{ 燭光} / 1 \text{ 燭光})=2$



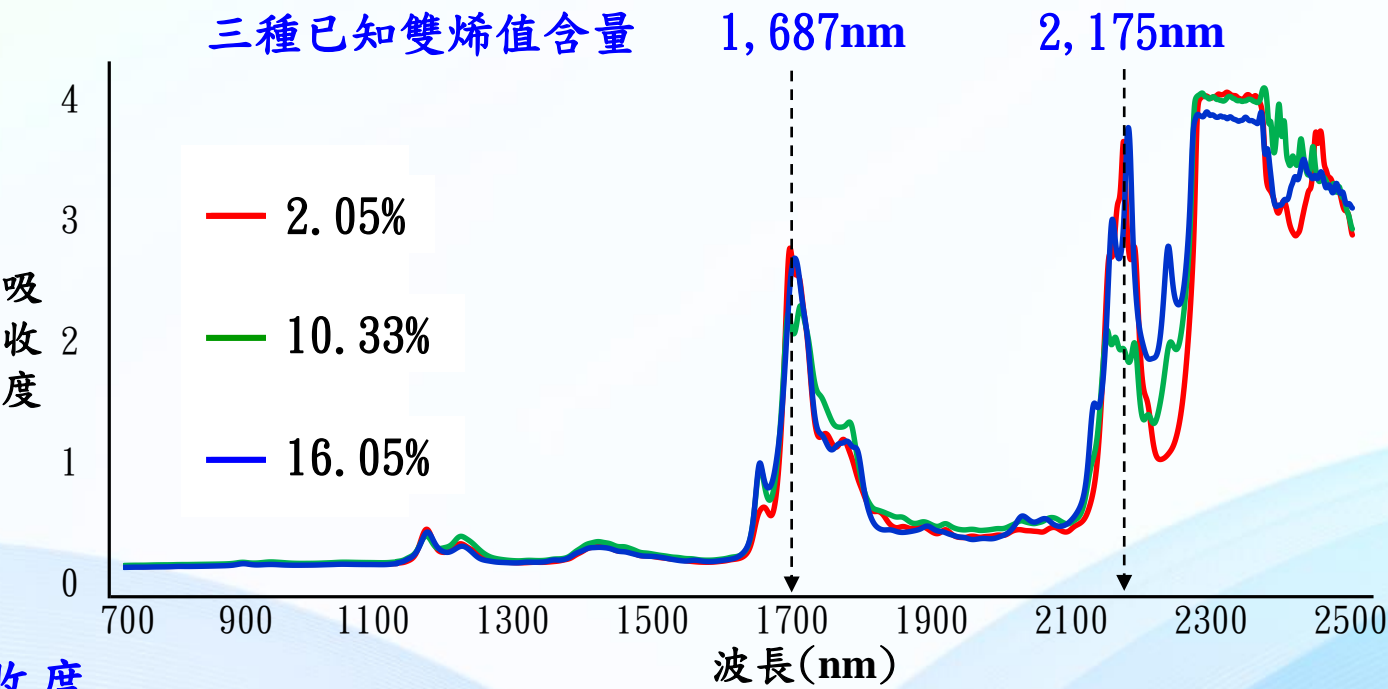
收集大量光譜圖建立AI模型

3.2 解決方案

➤ 舉例以三個不同濃度雙烯值樣品近紅外線光譜如下圖所示，在吸收度較明顯的波長段(下表以波長1,687nm及2,175nm為例)，濃度與吸收度的比值並無規則性，故無法使用一般光譜定量分析方法來反推雙烯值濃度，因此決定採AI演算法建立雙烯值判讀模型。

粗裂解汽油樣品光譜圖

波長 (nm)	濃度 (%)	吸收度	濃度/ 吸收度
1,687	2.05	2.47	0.83
	10.33	2.19	4.72
	16.05	2.63	6.10
2,175	2.05	2.70	0.76
	10.33	1.80	5.56
	16.05	3.51	4.57



註：上表相同波長，濃度/吸收度並無規則性，無法由人工從吸收度反推濃度。

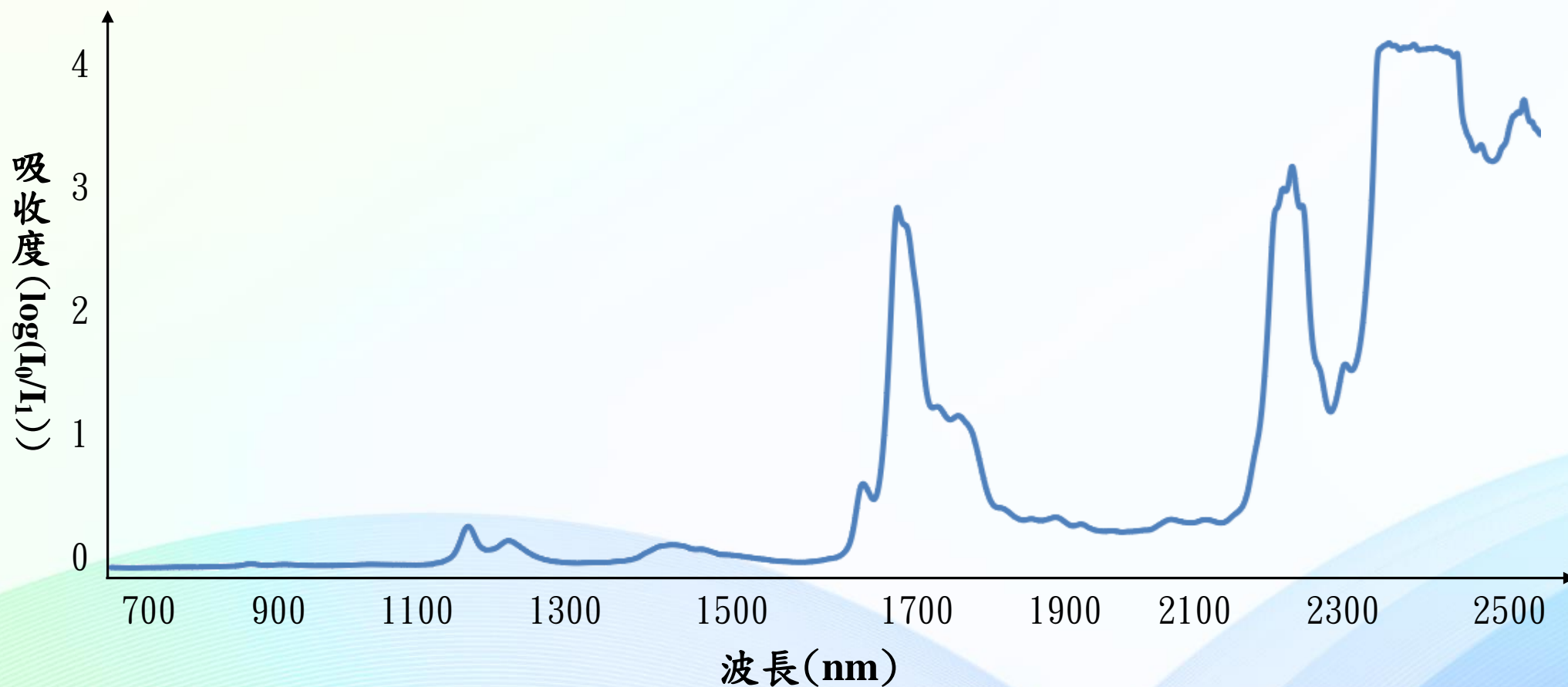
4. AI模型開發歷程



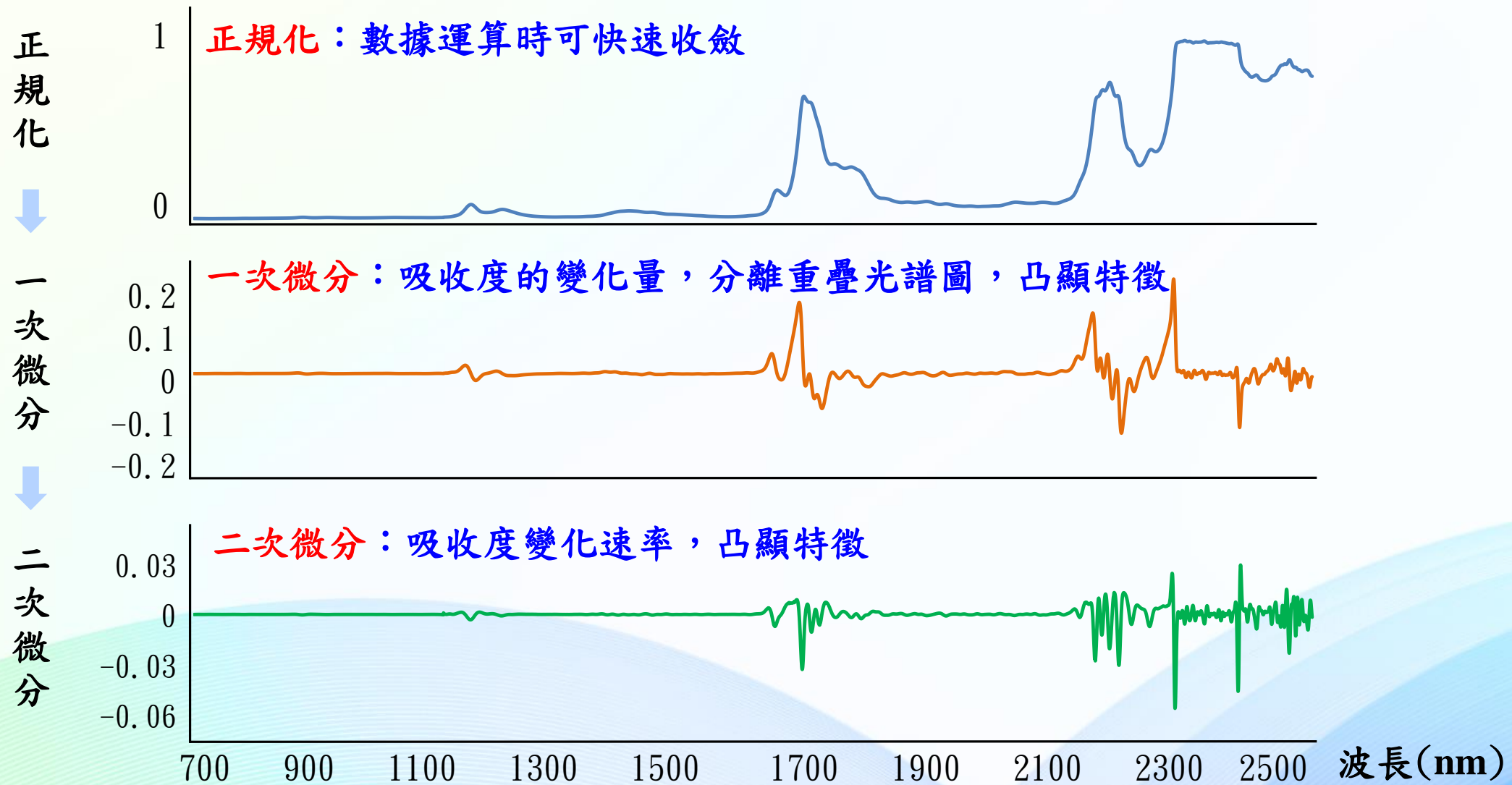
進程	第一階段	第二階段	第三階段	第四階段
項目	數據收集	資料前處理	模型建立	驗證上線
說明	收集1,000張光譜圖資料，其中800張為訓練資料，200張為驗證資料。	利用下列三種方法獲取特徵資料： 1. 正規化 2. 一次微分 3. 二次微分	1. 選擇六種演算法搭配資料前處理方法進行建模。 2. 其中以偏最小平方法(PLS)的預測值平均相對誤差<5%為最佳模型。	1. 2018/10/18~11/17與傳統方法進行比對，驗證模型符合要求。 2. 2018/11/18正式上線運作。 3. 定期以已知含量樣品驗證模型可靠性。

註：相對誤差= | 預測值-檢測值 | / 檢測值×100%

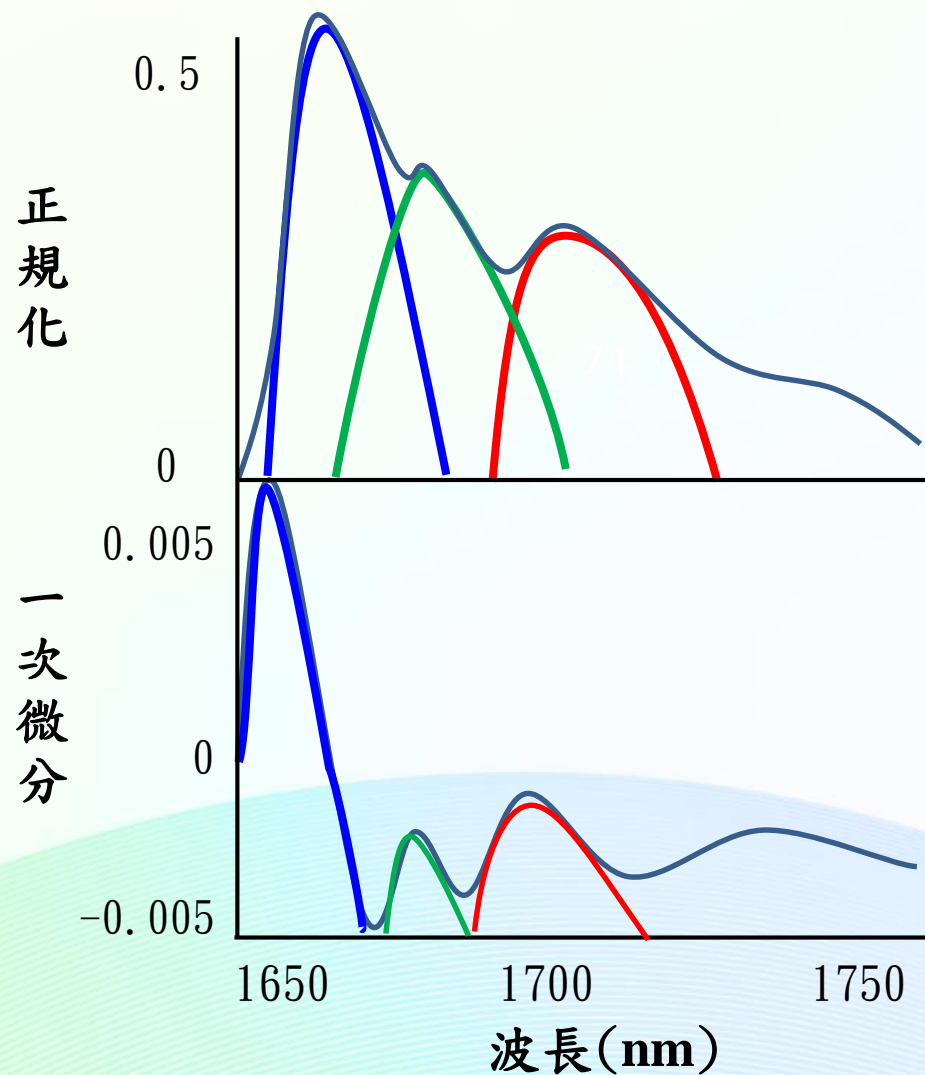
- 時間：2018年7月~ 2018年9月。
- 數量：1,000張吸收光譜圖。
- 收集範圍：每0.5nm記錄1個吸收強度，故每張光譜圖有3,600個，合計360萬個數據。



- 正規化：將數據範圍壓縮到0~1。
- 吸收光譜圖微分(一次或二次)：消除干擾，以凸顯特徵(分辨率)。



- 未微分時圖譜有部分重疊現象，採一次微分可以排除圖譜重疊現象，凸顯特徵。



未微分：圖譜有部分重疊現象。

一次微分：分離重疊光譜圖，凸顯特徵。

► 演算法選擇：

本案雙烯值預測模型選擇6種迴歸演算法，結果以DNN搭配一次微分獲得最小MAE=0.47。

資料前處理 \ 演算法		MLR	RIDGE	LASSO	PCR	PLS	DNN
正規化	未微分	1.51	1.51	0.80	5.93	0.52	0.48
	一次微分	1.57	1.57	0.68	5.6	0.5	0.47
	二次微分	1.56	1.56	0.74	4.03	0.48	0.48

模型	中文描述
MLR	多重線性迴歸
RIDGE	嶺迴歸
LASSO	套索算法
PCR	主要成分迴歸
PLS	偏最小平方法
DNN	深度神經網路

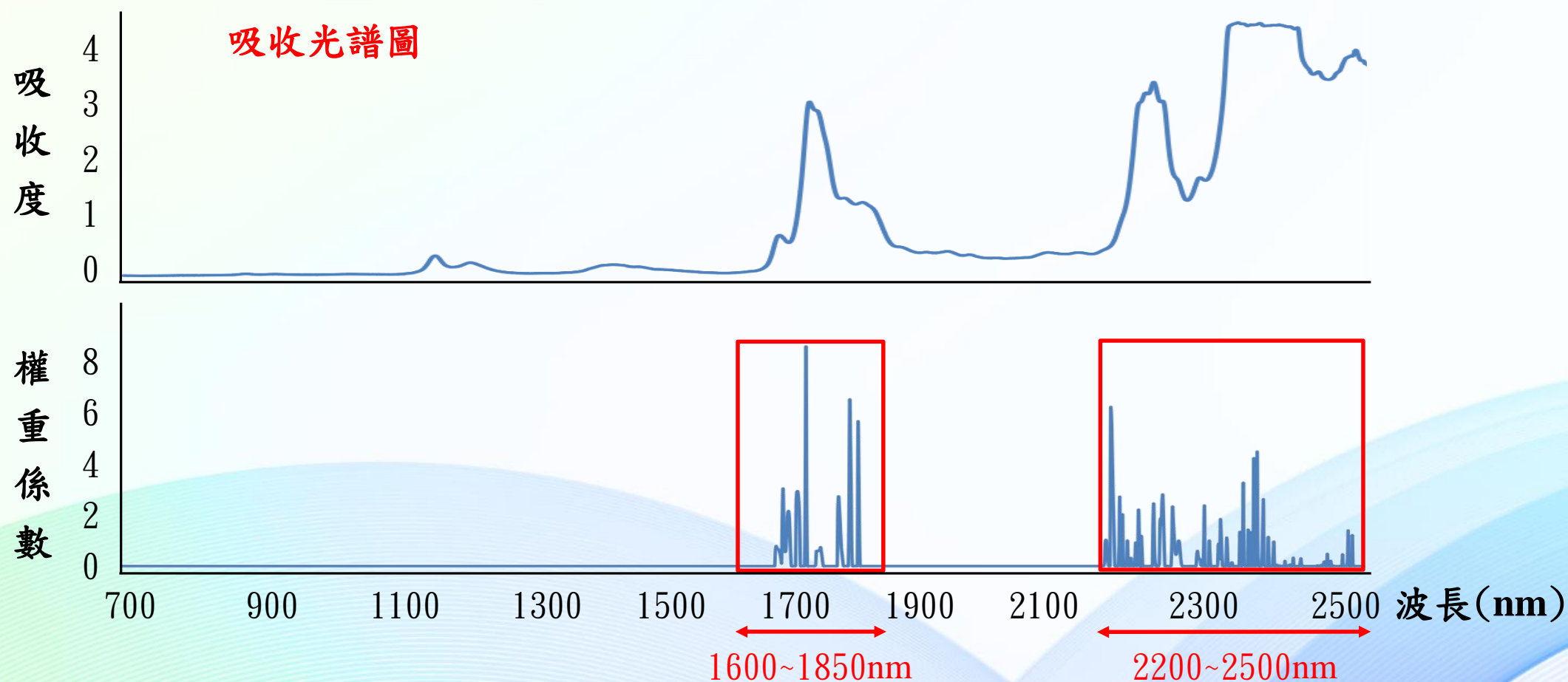
註：MAE為平均絕對誤差(Mean Absolute Error)

- 驗證平均相對誤差，氫化後雙烯值平均相對誤差為6.63%，無法符合目標< 5%，故需再進行AI模型優化。

資料前處理 \ 項目		DNN演算法 MAE值	氫化後雙烯值 平均相對誤差(%)
正規化	未微分	0.48	6.90%
	一次微分	0.47	6.63%
	二次微分	0.48	6.90%

➤ 特徵篩選：

為了進一步優化模型，降低MAE，對吸收光譜圖進行特徵權重係數分析。經評估具有明顯相關係數波長範圍為1,600~1,850nm及2,200~2,500nm兩區間，故選擇這兩區間數據重新模擬，特徵由3,600個降為1,100個。



➤ 模型優化：

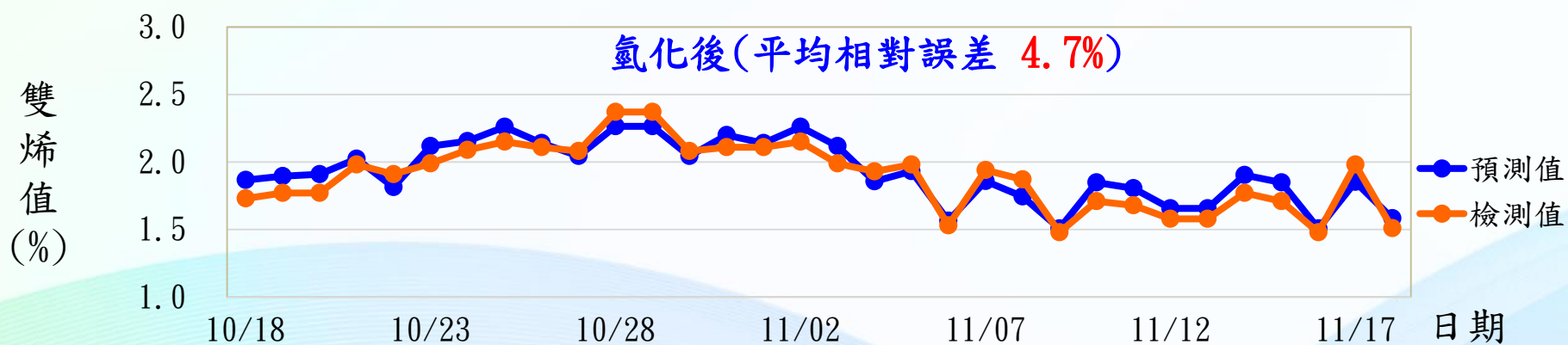
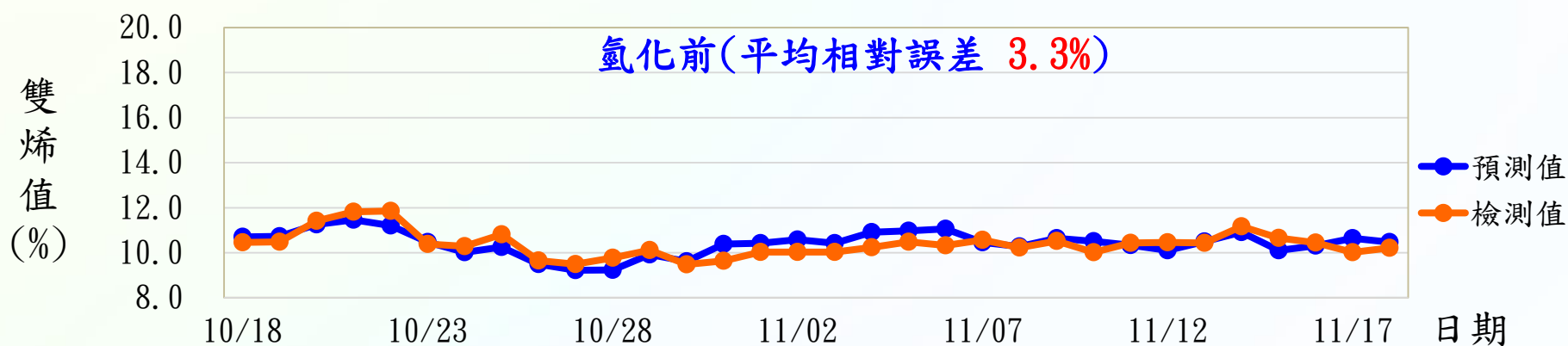
經挑選特定波長(1, 600~1, 850、2, 200~2, 500nm)的數據進行正規化及微分處理後，結果以**PLS**的**MAE**最小。

資料前處理 \ 演算法		MLR	RIDGE	LASSO	PCR	PLS	DNN
正規化	未微分	0.75	0.75	0.88	2.12	0.41	0.49
	一次微分	0.75	0.75	0.72	2.32	0.40	0.45
	二次微分	0.73	0.73	0.72	2.61	0.42	0.46

- 驗證平均相對誤差，未微分氫化後雙烯值平均相對誤差為5.02%，無法符合要求，所以再加上一次微分，氫化後可改善為4.7%，符合目標< 5%，故選擇PLS為本案的演算法。

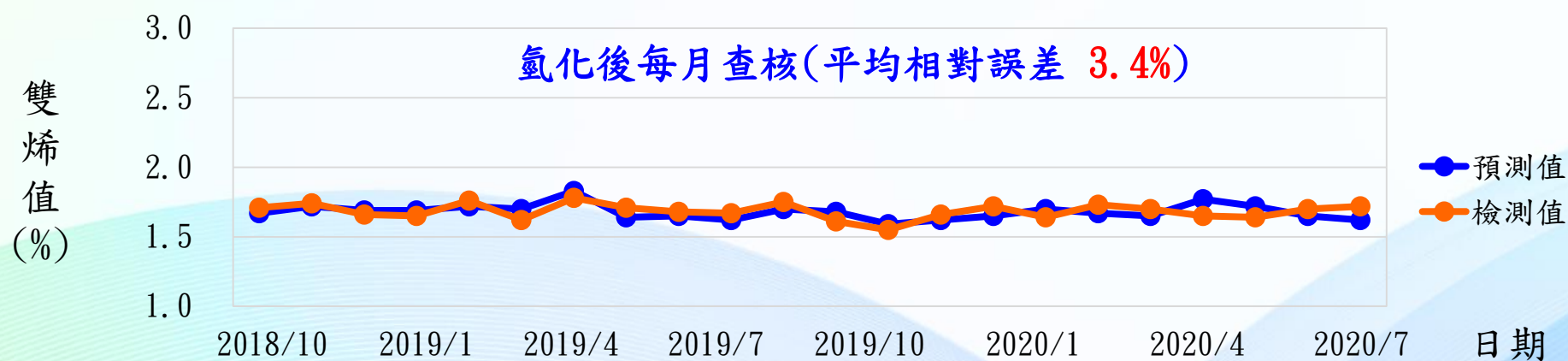
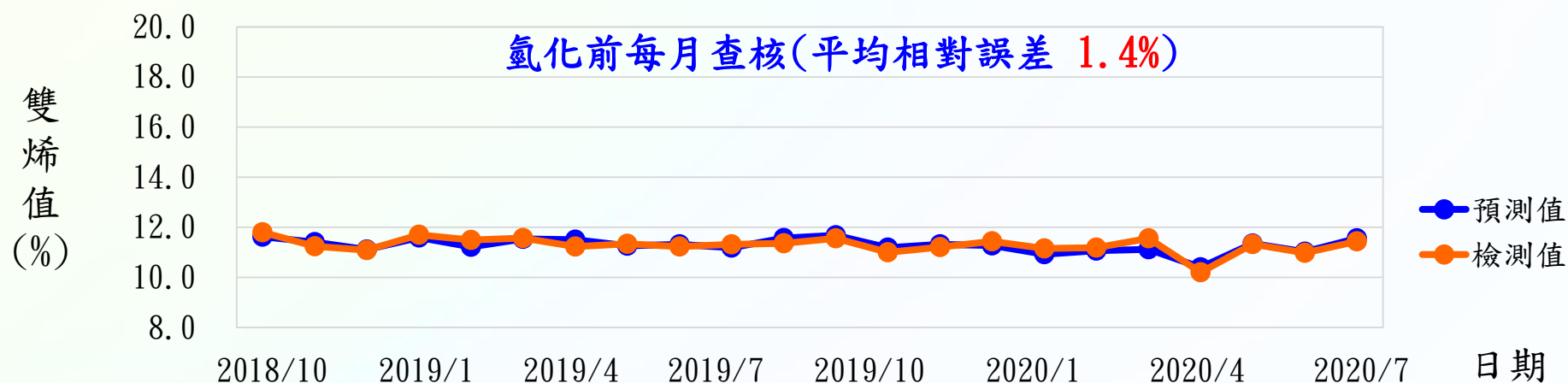
資料前處理 \ 項目		PLS演算法 MAE值	氫化後雙烯值 平均相對誤差(%)
正規化	未微分	0.41	5.02%
	一次微分	0.40	4.7%
	二次微分	0.42	5.3%

- 將PLS模型部署上線測試，驗證時間2018/10/18~11/17。
- 驗證結果：AI預測結果與實驗室數據相近，平均相對誤差 < 5%，符合使用需求。


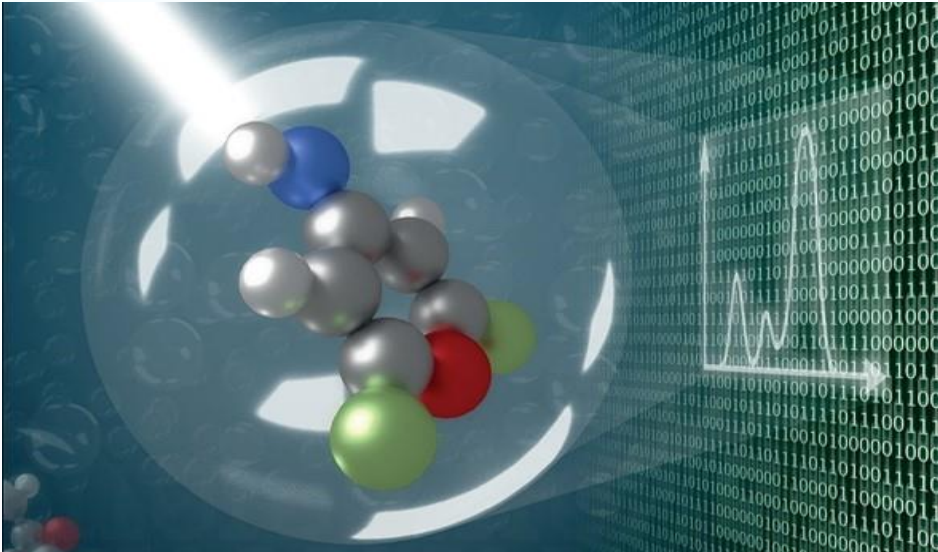


註：相對誤差 = $\frac{|\text{預測值} - \text{檢測值}|}{\text{檢測值}} \times 100\%$

- 模型部署上線後，每月持續進行AI預測值與實驗室檢驗值進行驗證。
- 驗證結果：2018年10月迄今，AI預測結果與實驗室數據平均相對誤差<5%，顯示模型長期穩定運作。



- 樣品檢驗時間可由5小時減至3分鐘。
- 雙烯值分析化學藥品使用量由295公升/年減至1.1公升/年(驗證用)。

改善前	改善後
	
檢驗時間5小時	檢驗時間3分鐘
需繁複樣品前處理	不需樣品前處理
滴定分析	光學檢測 + AI判讀

因為AI，沒有藥品也可以做檢驗

6. 效益說明

一、投資費用：3,500千元。

近紅外線光譜儀：3,500千元。

二、年效益：2,549千元。

1. 改善後檢驗時間由原259小時/月，降至20小時/月，工時節省239小時/月(約為1.5個人力)，節省的人力已轉任新成立的材料化學組，負責核磁共振儀等高階儀器操作，工時效益為：

$239\text{小時/月} \times 610\text{元/小時} \times 12\text{月/年} = 1,749\text{千元/年}$ 。

2. 藥品費用每月減少46,375元；廢液處理費用每月減少6,150元，效益合計630千元/年。

3. 設備維護費用每月減少14,126元，效益為170千元/年。

三、回收年限：1.37年。

7. 後續推動事項

- 以此技術平行展開共21案，預計年效益10,144千元，已完成11案，年效益5,651千元，改善中10案，預計2020年底完成。

項次	事業部	改善項目	訓練資料量 (張數)	年效益 (千元)	完成日
1	台塑聚烯部	LLDPE粉體密度分析	2,600	400	2019.03.20
2		EVA粒子中醋酸乙烯酯分析	500	50	2019.09.30
3		HDPE粒子密度分析	2,100	1,354	2020.03.31
4		HDPE粒子黏度分析	1,600	18	2020.08.20
5		LLDPE粒子熔融指數分析	700	18	2020.08.20
6	台塑塑膠部	PVC粉假比重分析	400	360	2020.01.31
7		PVC粉樣品聚合度分析	500	450	2020.01.31
8	台塑化品部	ECH廠樣品化學需氧量分析	197	100	2020.06.30
9	塑化烯烴部	油品雙烯值AI分析	1,000	2,549	2018.11.18
10		汽油氫化區樣品密度分析	2,686	176	2019.11.30
11		輕油雷氏蒸氣壓分析	328	176	2019.12.31
12	其他改善中項目共10案			4,493	(2020.12.30)
合計共21項				10,144	-

報 告 完 畢
恭 請 指 導

附件. 英文專有名詞資料表

英文名詞	英文全名	中文名稱	說明
MAE	<u>M</u> ean <u>A</u> bsolute <u>E</u> rror	平均絕對誤差	多筆測量的絕對誤差取絕對值後再求平均值。
MLR	<u>M</u> ultiple <u>L</u> inear <u>R</u> egression	多重線性迴歸	探討多個預測變數及一個依變數之間的關係。
RIDGE	RIDGE	嶺迴歸	一種同時進行特徵選擇和平方正則化的回歸分析方法。
LASSO	<u>L</u> east <u>A</u> bsolute <u>S</u> hrinkage and <u>S</u> election <u>O</u> perator	套索算法	一種同時進行特徵選擇和絕對值正則化的回歸分析方法。
PCR	<u>P</u> rincipal <u>C</u> omponents <u>R</u> egression	主要成分迴歸	將變數的高維空間投影至特徵向量的低維空間，進而降低變數維度並能維持原有的資訊量。
PLS	<u>P</u> artial <u>L</u> east <u>S</u> quares	偏最小平方法	通過投影預測變數和觀測變數到一個新空間來尋找一個線性迴歸模型。
DNN	<u>D</u> eep <u>N</u> eural <u>N</u> etworks	深度神經網路	以神經網路為基礎架構，藉由加深隱藏層來取代傳統的特徵工程，進而提效能。