

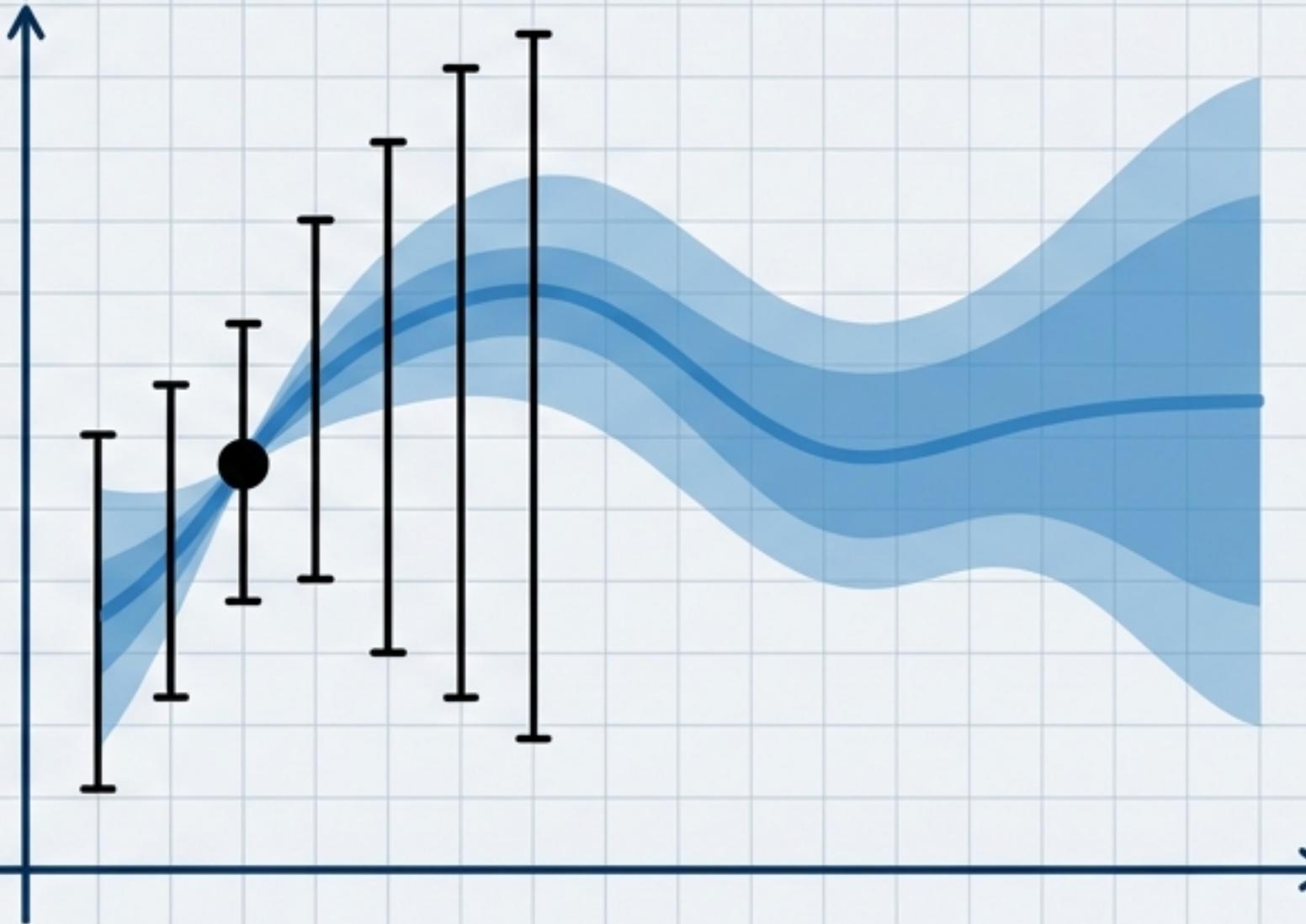


# Unit 11: 高斯過程回歸

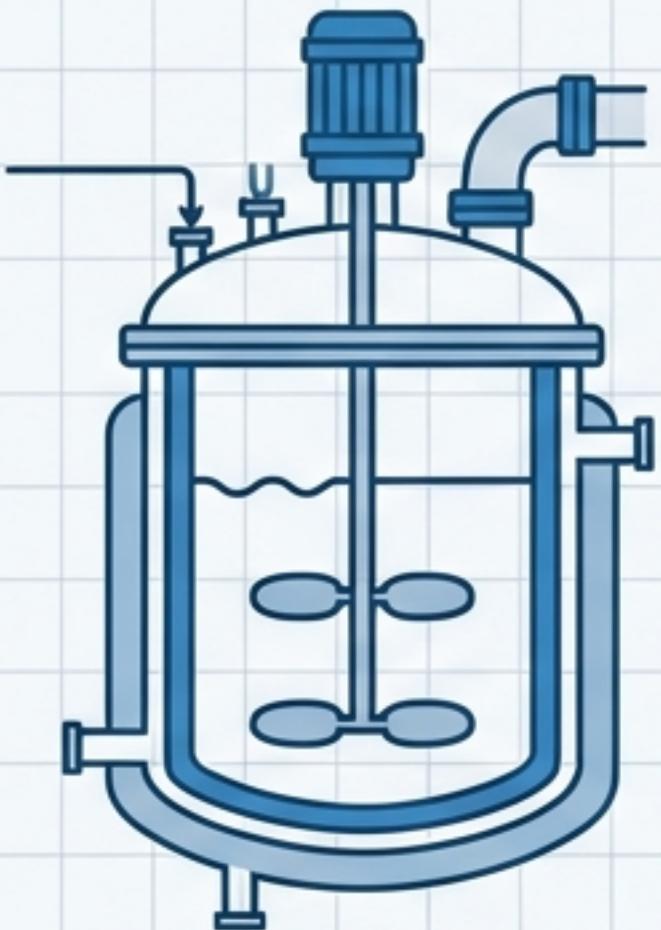
# Gaussian Process Regression

專為小樣本與高成本實驗設計的機率模型

Project Specification	
COURSE CODE	: CHE-AI-114
INSTRUCTOR	: 莊曜禎 助理教授
DEPARTMENT	: 逢甲大學 化工系 智慧程序系統工程實驗室
DATE	: 2026-01-28



# 化工現實困境：昂貴實驗與小數據

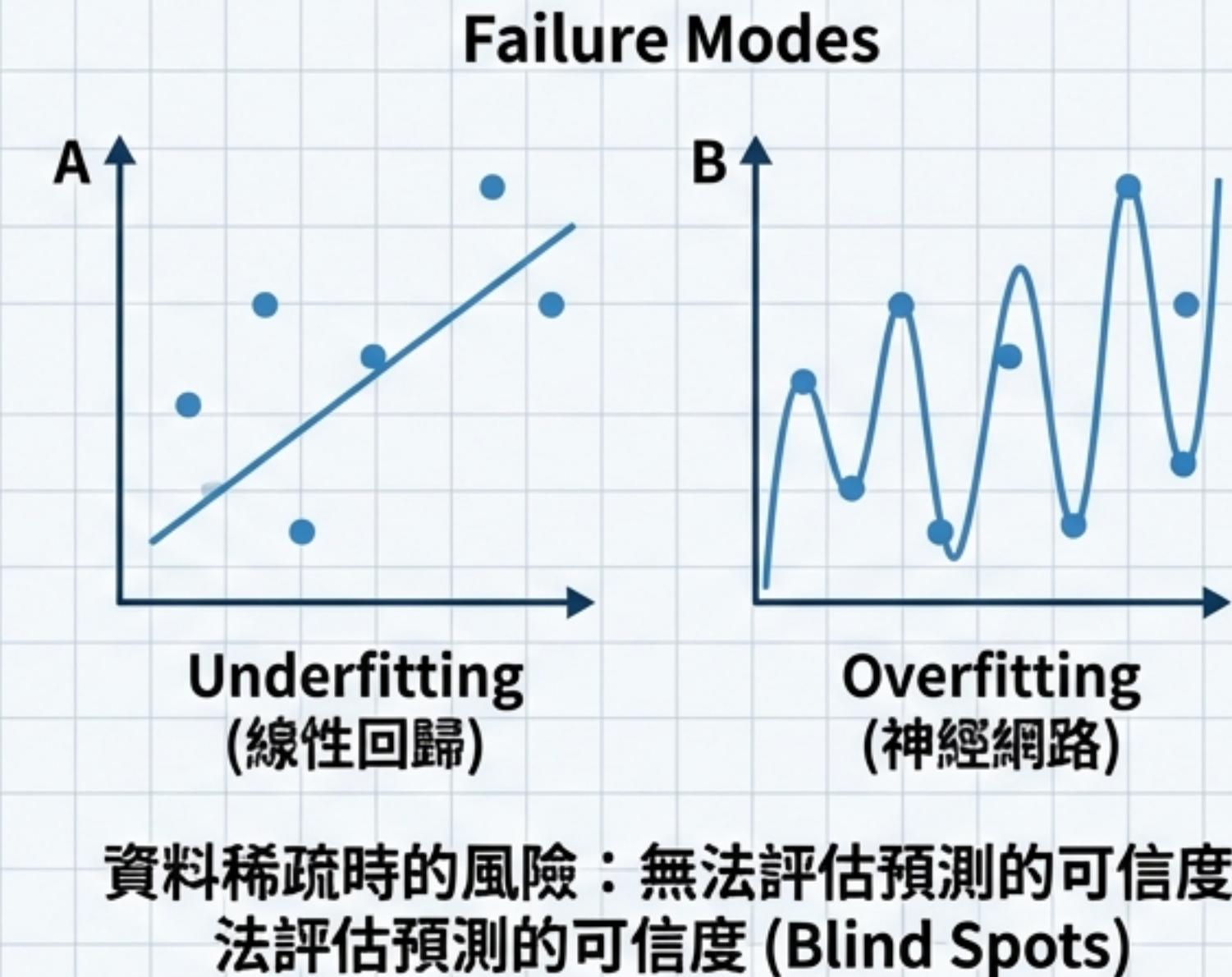


傳統 AI 需要海量數據，  
但化工實驗往往受限：

成本高昂：  
NT\$ 10,000 /  
每個數據點

耗時：  
2 天 / 每次實驗

資源限制：  
樣品製備困難



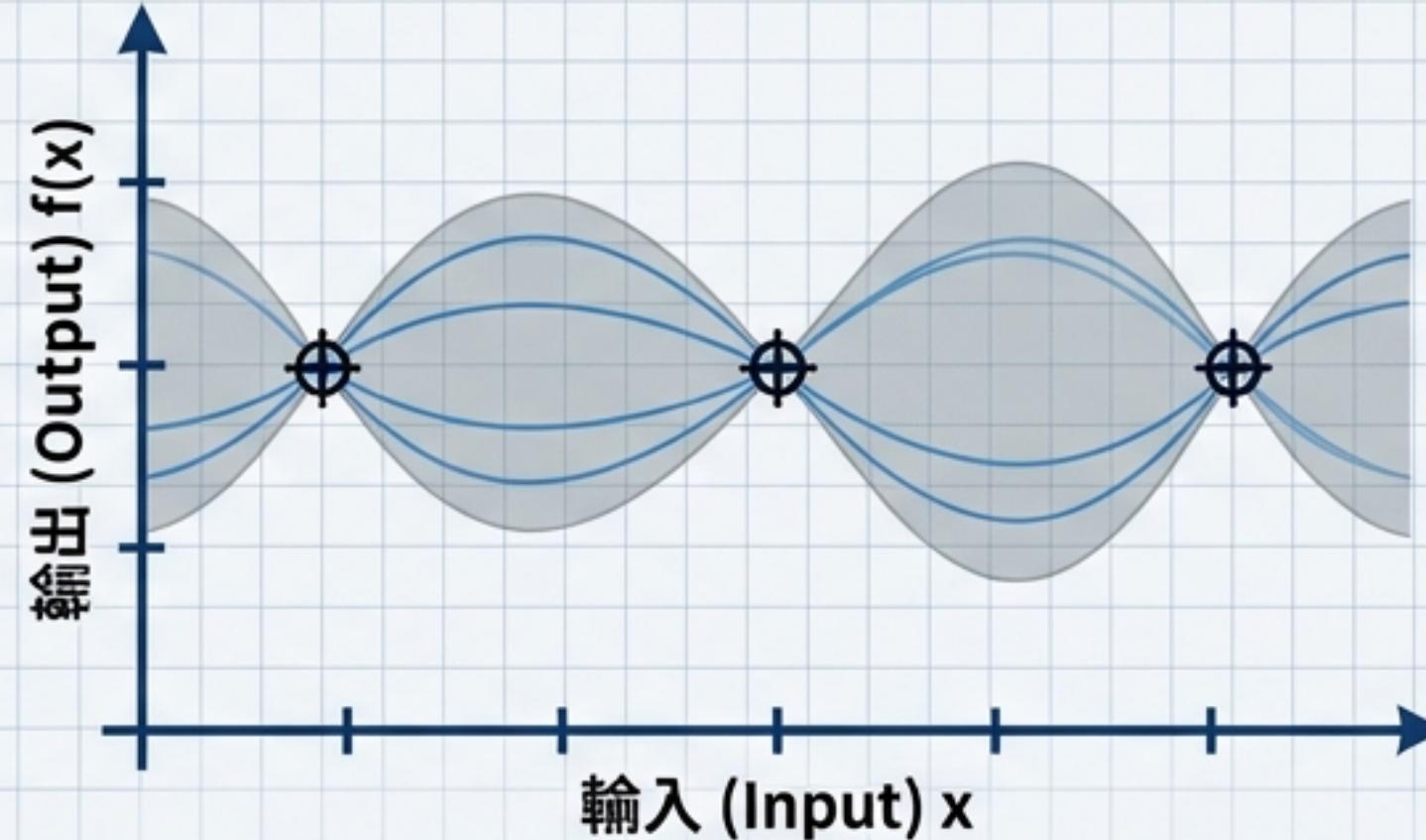
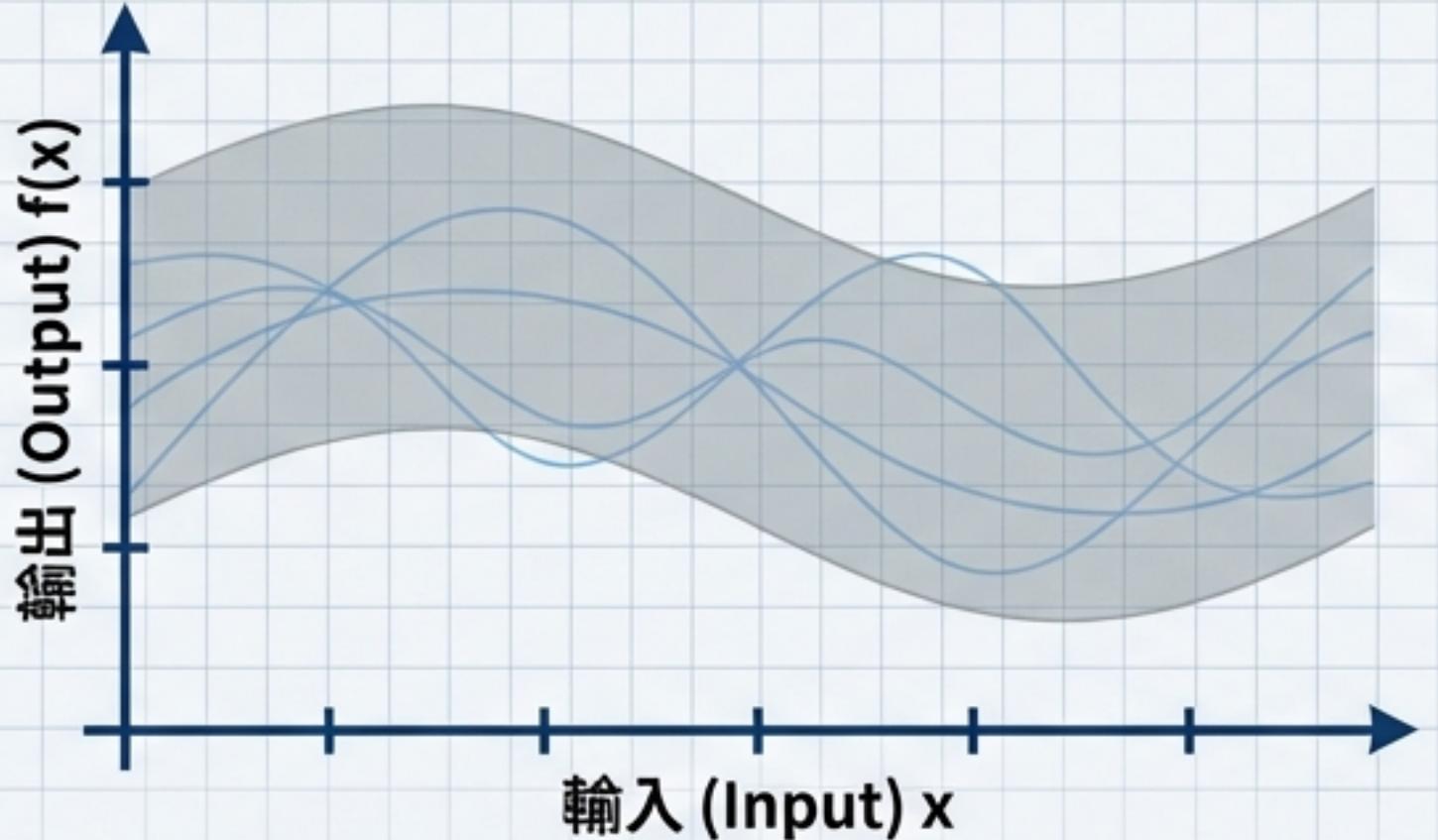
**SOLUTION : 需要一個能理解『自身無知』的模型 — GPR**

# 高斯過程 (GPR) 的核心概念

不只是擬合一條線，而是對函數本身建立機率分佈

先驗分佈 (Prior) - 初始信念

後驗分佈 (Posterior) - 觀測後更新



$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

由 均值函數  $m(x)$  與 核函數  $k(x, x')$  定義

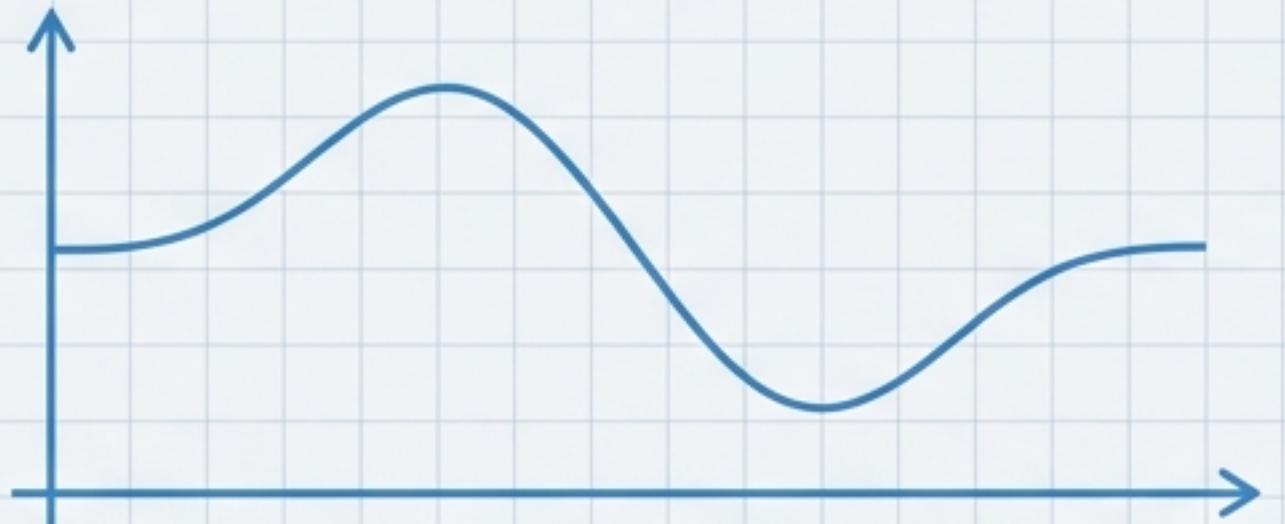
想像它是一個無限維度的多變量高斯分佈。

# 核函數 (Kernels)：模型的「透鏡」

定義數據點之間的相似性與函數特性

## RBF Kernel (Squared Exponential)

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right)$$



無限可微，極度平滑。適用於理想化系統。

## Matérn Kernel (Matérn 5/2)

$$k(r) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \dots K_\nu(\dots)$$

$r = \|\mathbf{x} - \mathbf{x}'\|/\ell$



有限可微，允許局部粗糙度。更適合真實化物理系統。

### 關鍵參數 Key Parameters:

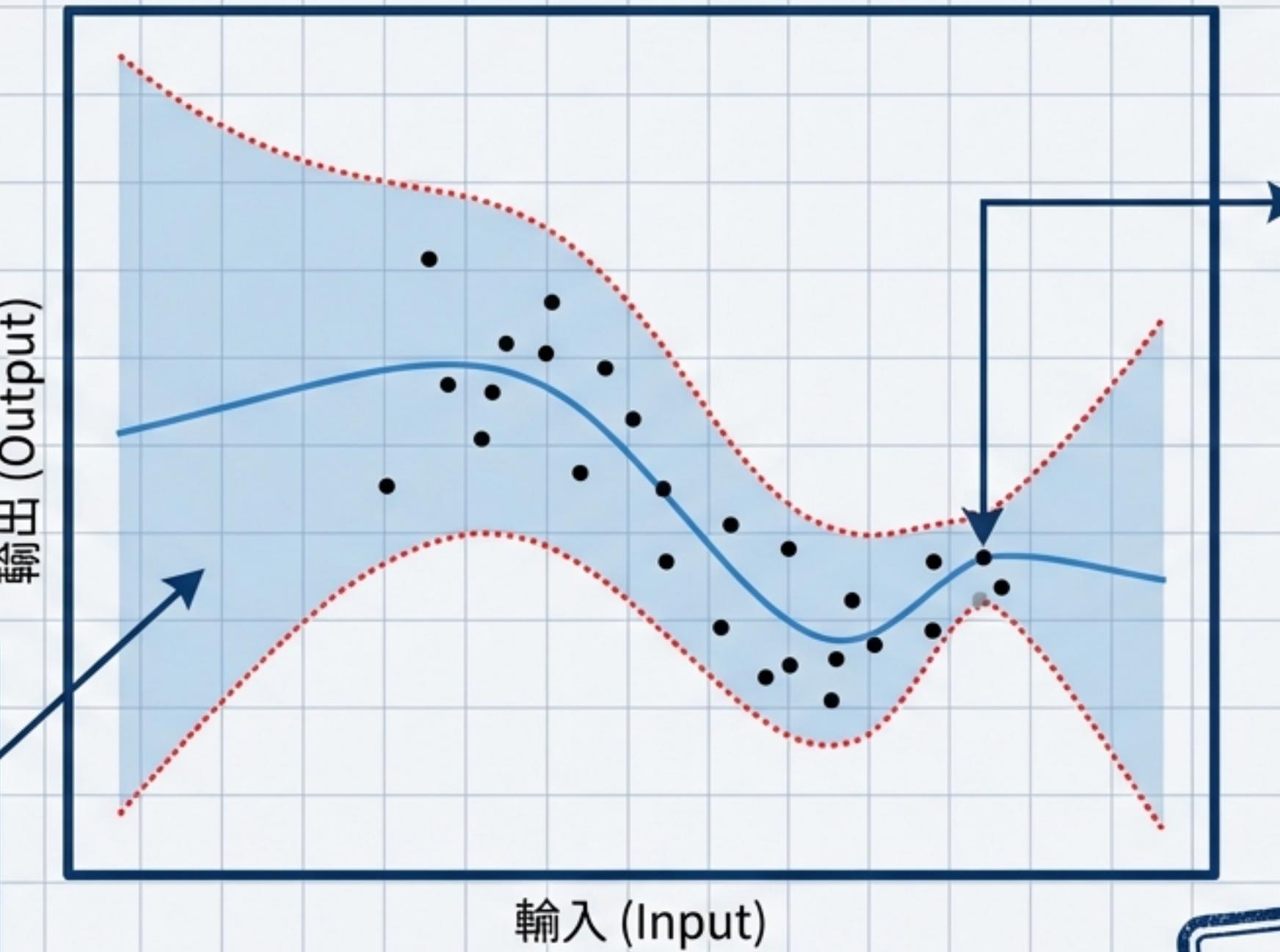
- Length Scale ( $\ell$ ) : 影響範圍 (Influence Range)
- Signal Variance ( $\sigma_f^2$ ): 波動幅度 (Amplitude)

TECHNICAL SPECIFICATION

# 不確定性量化：工程師的安全邊界

## 1. 認知不確定性 (Epistemic Uncertainty)

- 我們不知道的事  
(What we don't know)
- 原因: 缺乏數據
- 對策: 增加實驗可消除



## 2. 隨機不確定性 (Aleatoric Uncertainty)

- 無法消除的噪聲
- 原因: 儀器誤差 ( $\sigma_n^2$ )

$$95\% \text{ CI} = \mu(\mathbf{x}_*) \pm 1.96\sigma(\mathbf{x}_*)$$

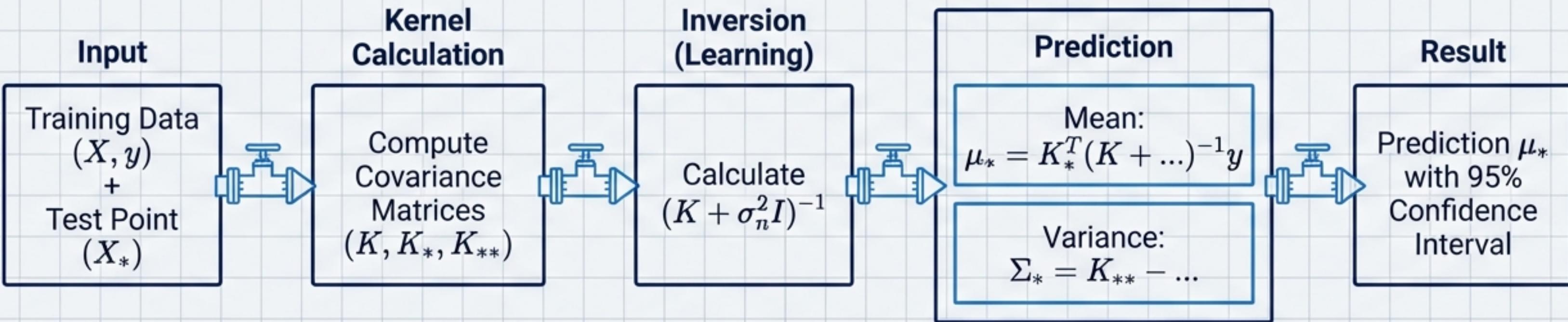
用途: 風險評估

# 模型規格比較表 (Model Specs)

特性 (Feature)	線性回歸 (Linear)	隨機森林 (RF)	神經網路 (NN)	GPR (本單元)
小樣本性能 (Small Data)	Low	Medium	Poor	Excellent (✓✓✓)
不確定性量化 (Uncertainty)	Limited	Statistical	No	Bayesian (Native)
非線性建模 (Non-linear)	No	Yes	Yes	Yes
計算成本 (Compute Cost)	Very Low	Low	High	High ( $O(N^3)$ )

GPR 是唯一能同時提供『高非線性擬合』與『嚴格不確定性量化』的小樣本工具。

# 運算流程圖：從先驗到預測



# 實戰案例：觸媒反應優化

## Case Study: Catalyst Reaction Optimization

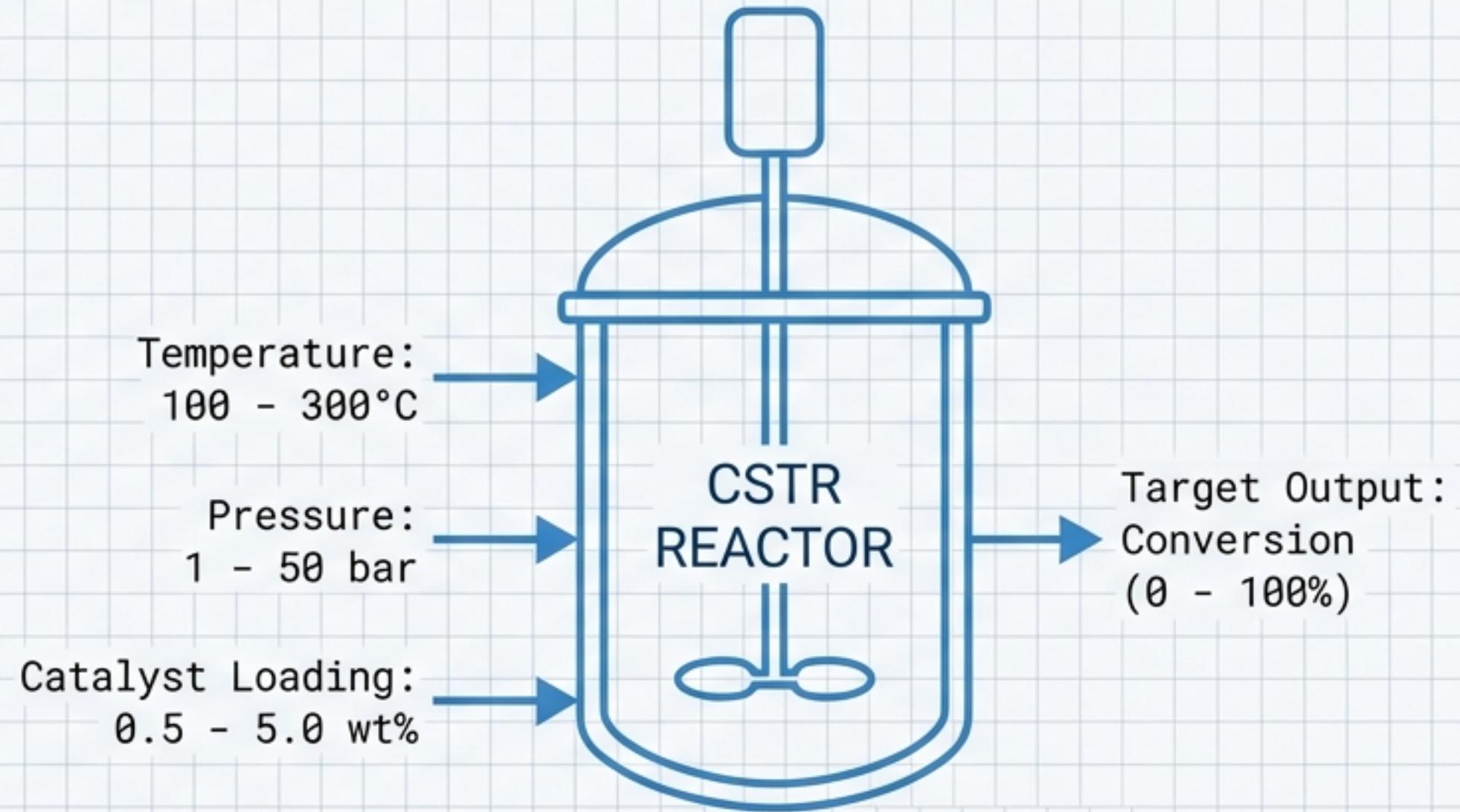
### Mission Profile

### Mission Objective

尋找最佳操作條件以最大化轉化率  
(Conversion)。

### Constraint

實驗昂貴且費時。必須以最少實驗  
次數達成目標。



DATASET: 100 Samples  
(Latin Hypercube Sampling)

# 程式實作：建構 GPR 控制器

```
1 from sklearn.gaussian_process import GaussianProcessRegressor  
2 from sklearn.gaussian_process.kernels import RBF, Matern  
3  
4 # 1. 定義核函數 (The Engine)  
5 kernel = 1.0 * Matern(length_scale=1.0, nu=2.5) ← Matérn 5/2 setting  
6  
7 # 2. 初始化模型 (Controller Setup)  
8 gpr = GaussianProcessRegressor(  
9     kernel=kernel,  
10    alpha=0.32, ← Handles noise level (Noise Level) ← experimental  
11    n_restarts_optimizer=15, # 避免局部最佳解  
12    normalize_y=True # 標準化目標變數  
13 )  
14  
15 # 3. 訓練 (Calibration)  
16 gpr.fit(X_train, y_train)
```

# 核函數選擇與校正 (Kernel Selection)

為什麼物理特性很重要? (Why Physics Matters)

	Kernel	R <sup>2</sup>	Description
✓	Matérn 5/2	0.7281	允許二階可微，符合真實化學反應的非平滑特性。
	RBF	0.7170	假設無限平滑，對於複雜反應過於理想化。
	Matérn 3/2	0.7166	過於粗糙 (Too Rough)。

## Optimized Hyperparameters

Length Scale  $\ell \approx 1.54$

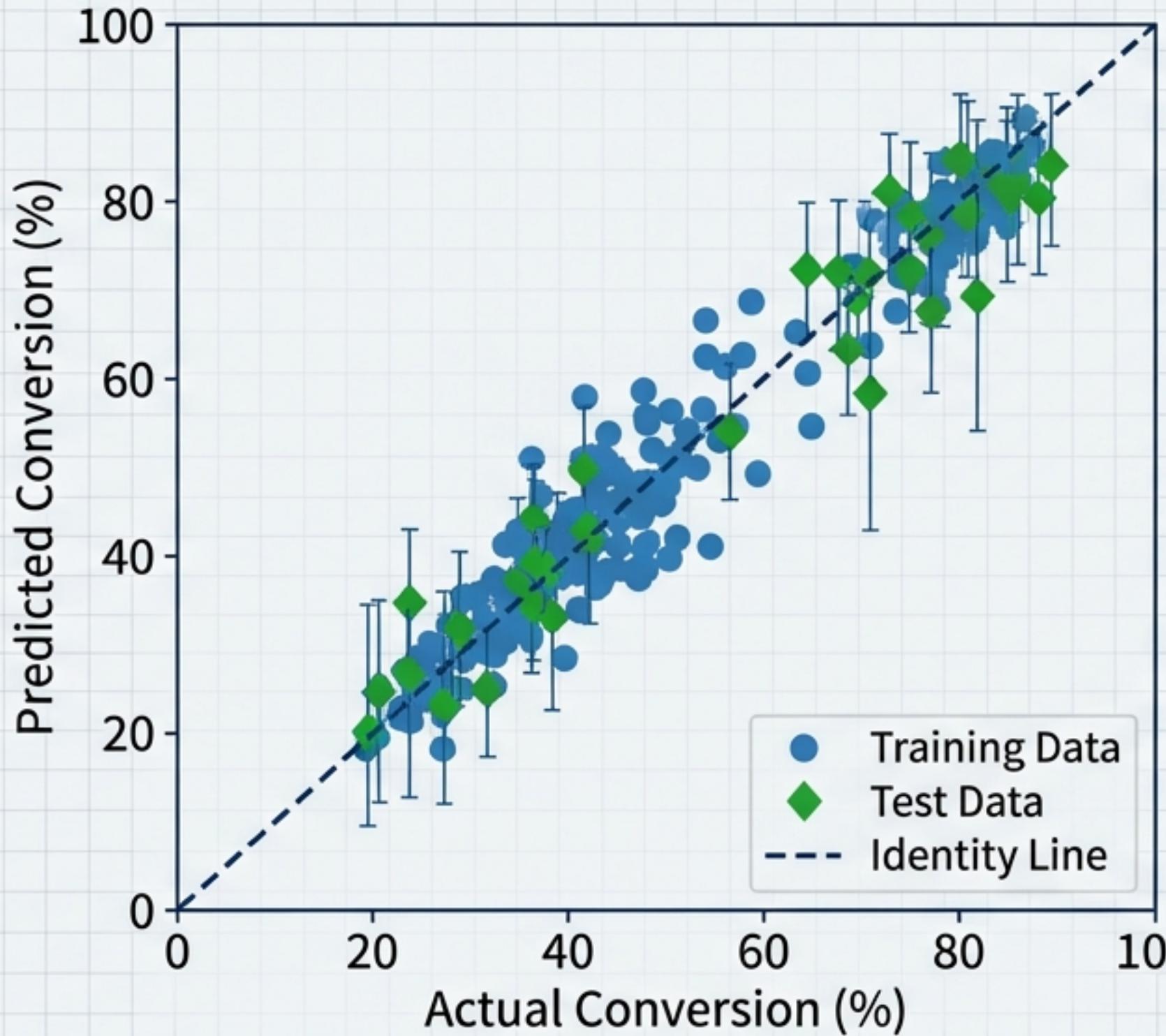
對應 30°C 溫差

Noise Level  $\alpha = 0.32$

對應 RMSE ~7%

Matérn 5/2 提供了最佳的預測精度與物理意義。

# 預測結果可視化 (Visualization of Results)



## Results Summary

Test  $R^2: 0.72$

RMSE: 7.0% (Acceptable)

CI Coverage: 88%

模型準確捕捉了高轉化率區域 ( $>90\%$ ) 的趨勢，且信賴區間合理反映了誤差範圍。

# 下一步：主動學習與序貫優化

利用「不確定性」選擇下一個實驗 (Active Learning)

Exploitation (利用)



選預測值最高點。

Exploration (探索)



選不確定性最高點。

UCB (Upper Confidence Bound)

$$\mu(x) + \kappa\sigma(x)$$

平衡兩者 (Balance)。

Current Model

$\overline{\mu(x)}$  Calculate UCB

Find Max UCB

Update Model

Run Experiment

相較隨機嘗試，  
可減少 30-50%  
的實驗次數。

# 操作限制與警示 (Operational Limits)



## 計算瓶頸 (Computational Bottleneck)

複雜度  $O(N^3)$  (Matrix Inversion)

當數據點  $N > 10,000$  時，訓練時間將急劇增加。



Recommendation: Use Sparse GP for Big Data.

## 外推能力 (Extrapolation)

Mean-Reverting Behavior

GPR 在遠離訓練數據時，預測會回歸均值。



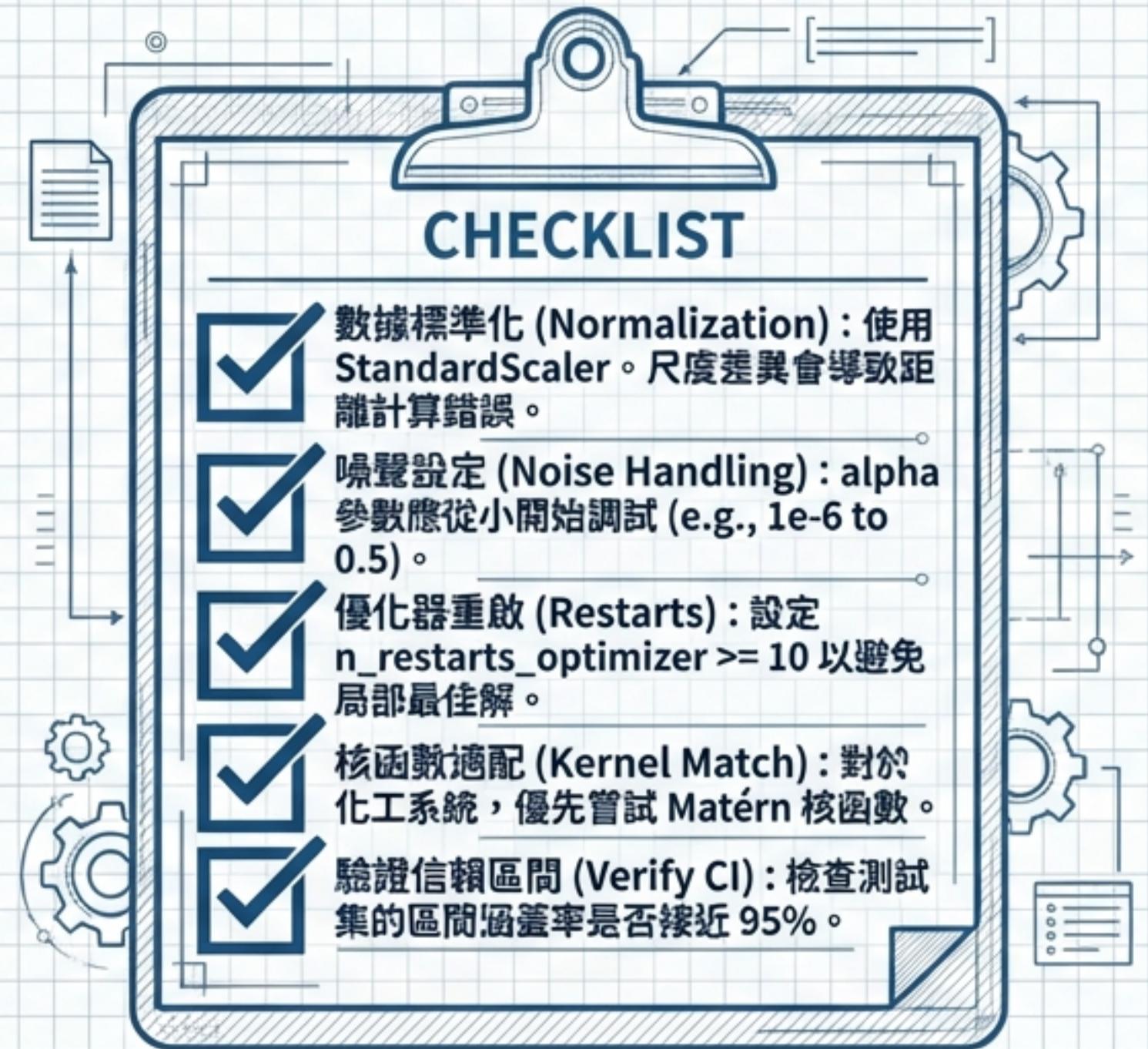
Warning: Do NOT trust predictions far outside experimental range ( $T > 300^\circ C$ ).

## 異常值敏感 (Outlier Sensitivity)

Gaussian Assumption

對極端異常值較敏感，需先剔除 Outliers。

# 最佳實踐檢查清單 (Best Practices)



*“Rigorous calibration ensures safety.”*

# 總結：您的數位化研發工具



## The Tool

GPR = 小樣本 + 不確定性量化  
(Small Data + Uncertainty)



## The Application

適用於昂貴實驗、反應優化、  
風險評估。



## Action

實作: Unit11\_Gaussian\_Process\_Regression.ipynb

「在充滿不確定性的化工製程中，GPR 提供了我們需要的導航地圖。」