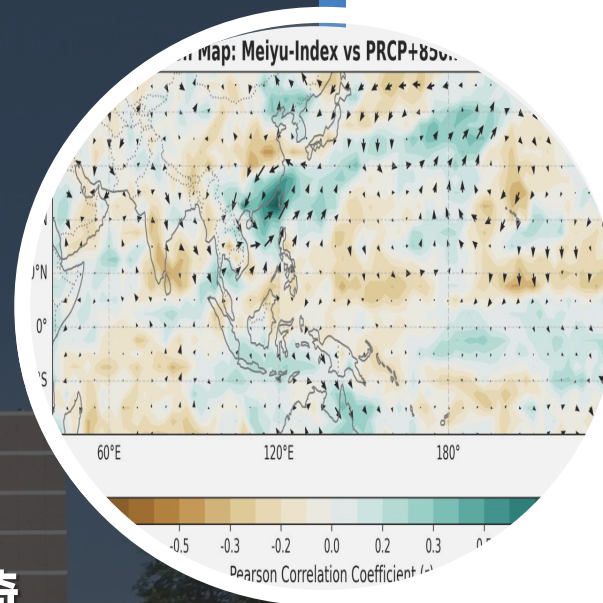




結合 VAE 與 XGBoost 開發梅雨季臺灣雨量預報產品

李清騰、林昀靜、陳昀靖、羅資婷、潘琦
中央氣象署/海象氣候組

115年第四十屆天氣分析與預報研討會
2026/09/02



臺灣氣候監測

開啟即時監測畫面 >>

2026年7月回顧

官方預報

聖嬰展望

氣候模式預報

全球

東亞

氣候主題預報

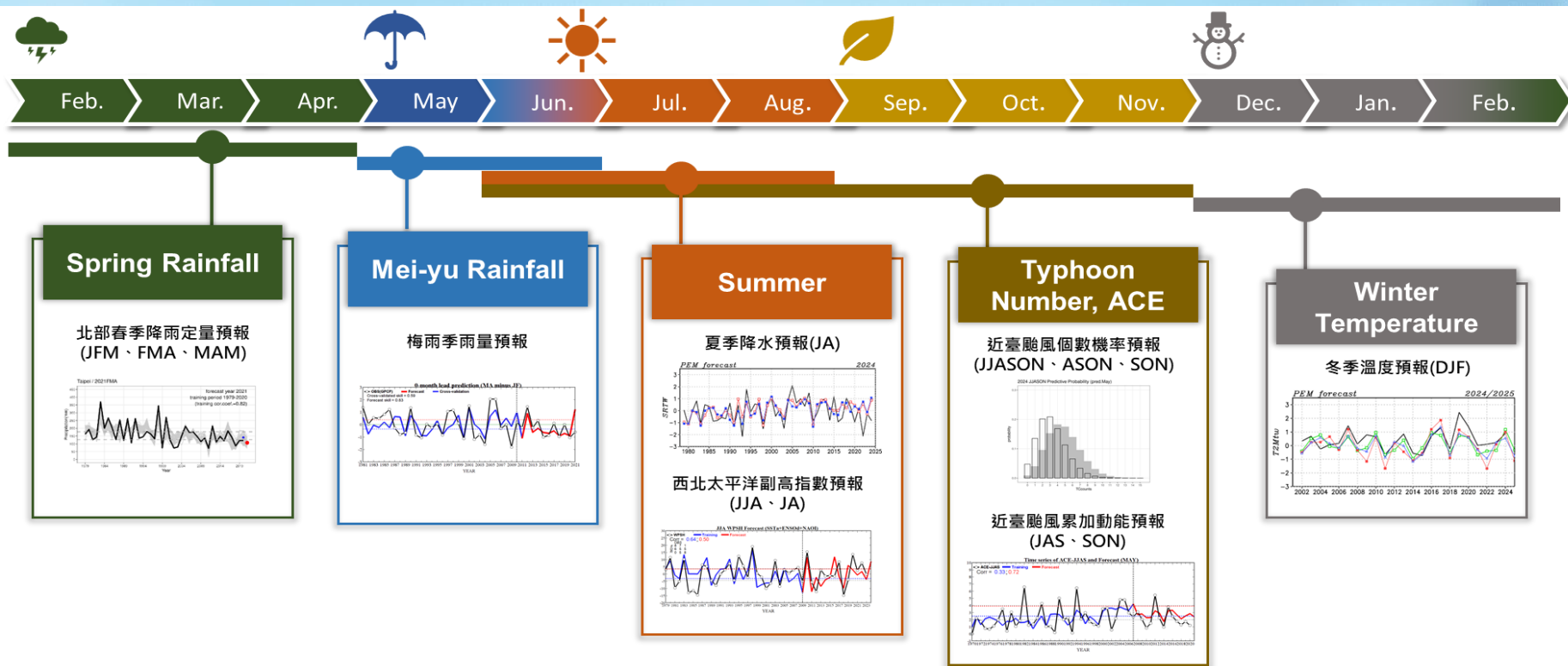
冬季

春季

颱風季



物理概念統計模型 (Physical Empirical model)



1. 資料: NCEP CFSR Data (T2m) & ERSSTv5 (SST) and climatology is 1981-2010
2. 預報對象: GPCP (area mean over 119E-123E, 21N-26N)
3. 預報因子 : WNPT(SST), IOWPT (SST), NAT(SST), EAT(T2m)
4. 統計迴歸模式:

Lead0-month: $TWRI = -0.531 * WNPT + 0.499 * NAT + 0.291 * EAT$ (ini: Apr)

Lead1-month: $TWRI = 0.393 * IOWPT + 0.374 * NAT + 0.329 * EAT$ (ini: Mar)

Lead2-month: $TWRI = 0.406 * IOWPT + 0.38 * NAT + 0.147 * EAT$ (ini: Feb)

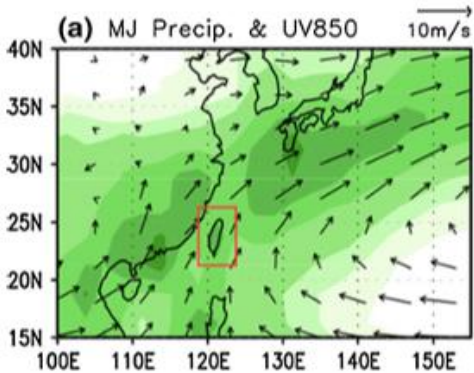


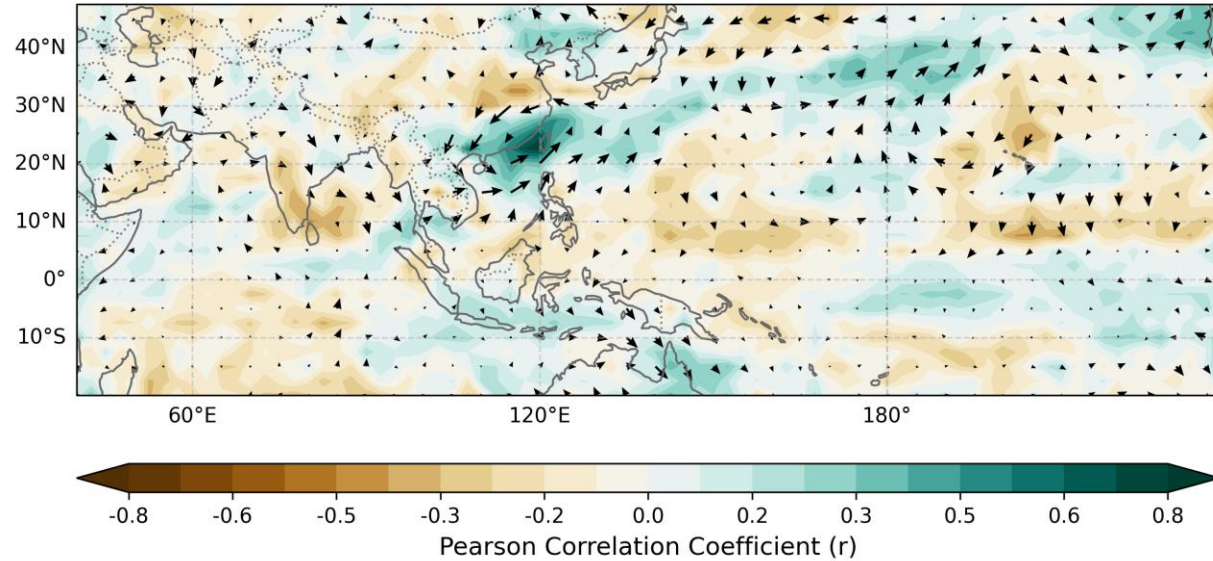
Table 1 Definition of each of 0-, 1-, and 2-month lead predictors selected for the prediction of MJ Taiwan precipitation variability

Predictor	Meaning	Definition		
		0-month lead (MA-minus-JF)	1-month lead (Mar-minus-Jan)	2-month lead (Feb-minus-Dec)
WNPT	Western North Pacific SST tendency	15N–25N, 130E–180E	–	–
IOWPT	Indo-Pacific warm pool SST tendency	–	(10S–10N, 80E–120E) – (10N–20N, 125E–160E)	(5S–15N, 70E–100E) + (15N–25N, 125E–160E) + (5S–5N, 170E–160W) – (5S–10N, 125E–160E)
NAT	Tripolar North Atlantic SST tendency	(0–20N, 60W–15W) + (45N–55N, 55W–40W) – (25N–40N, 65W–40W)	(0–20N, 60W–15W) + (45N–55N, 55W–40W) – (25N–40N, 65W–40W)	(0–15N, 60W–20W) + (35N–50N, 70W–30W) – (20N–30N, 80W–50W)
EAT	East Asia T2M tendency	30N–60N, 110E–140E	40N–60N, 70E–130E	55N–75N, 90E–140E

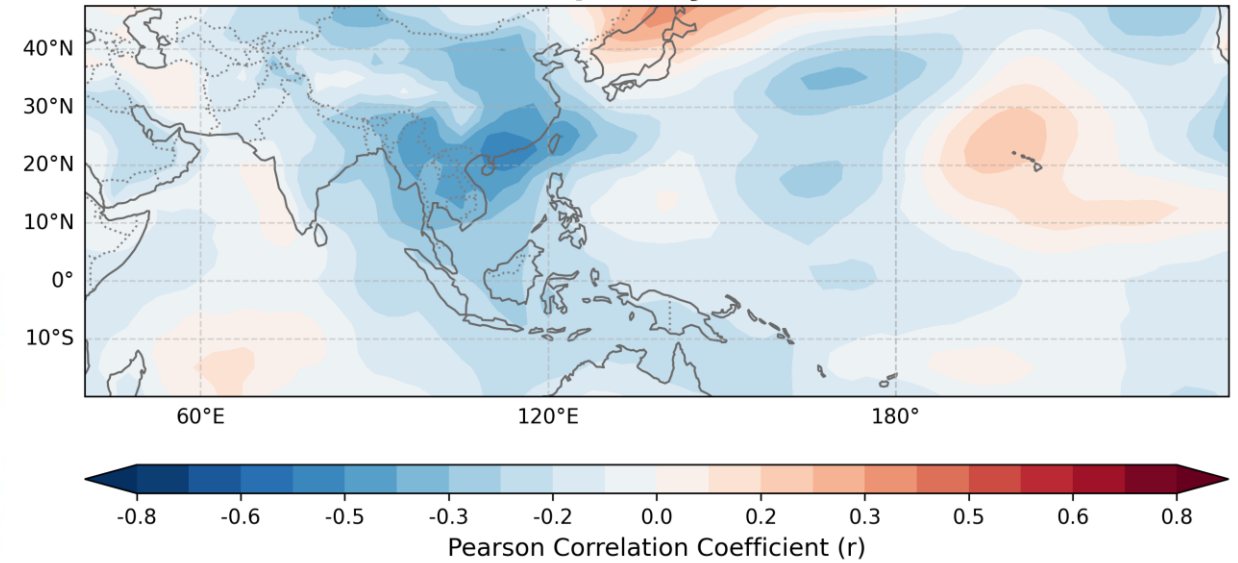


Correlation Map Between Meiyu-Index and MJ-Variables

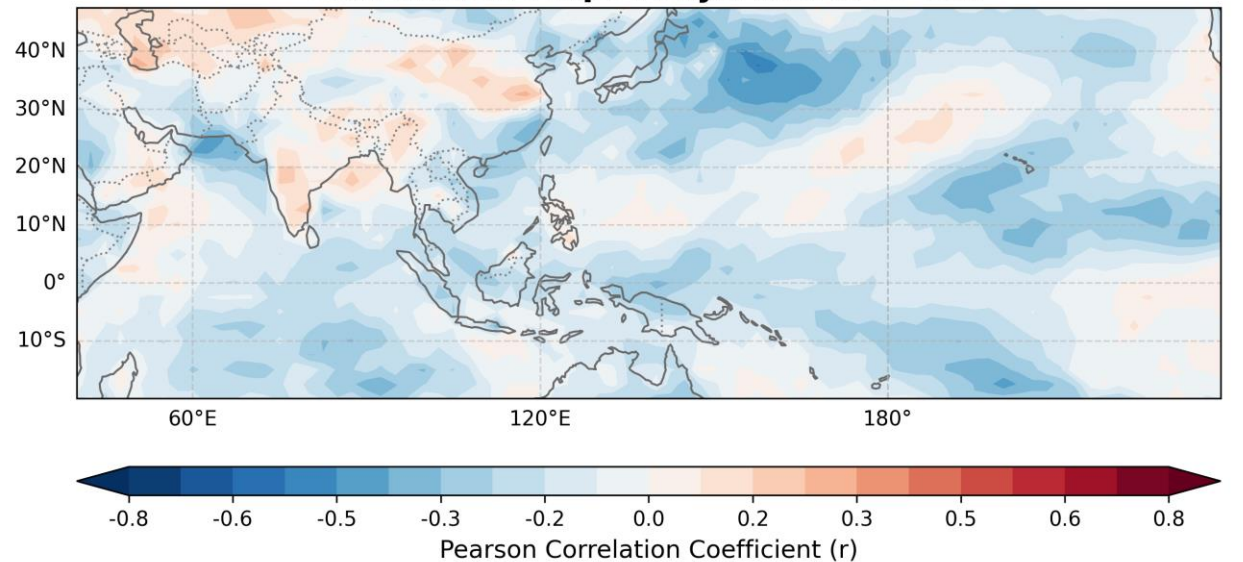
Correlation Map: Meiyu-Index vs PRCP+850hPa Wind (MJ)



Correlation Map: Meiyu-Index vs MSLP



Correlation Map: Meiyu-Index vs SST



歷史資料分析

預測因子：
ERA5/ NCEP CFSR
(過去歷史重分析資料)

預測目標：
臺灣梅雨季雨量指標
(歷史資料)

EOF分析
Correlation 分析
Regression 校驗

依據物理概念選
取合適預報因子

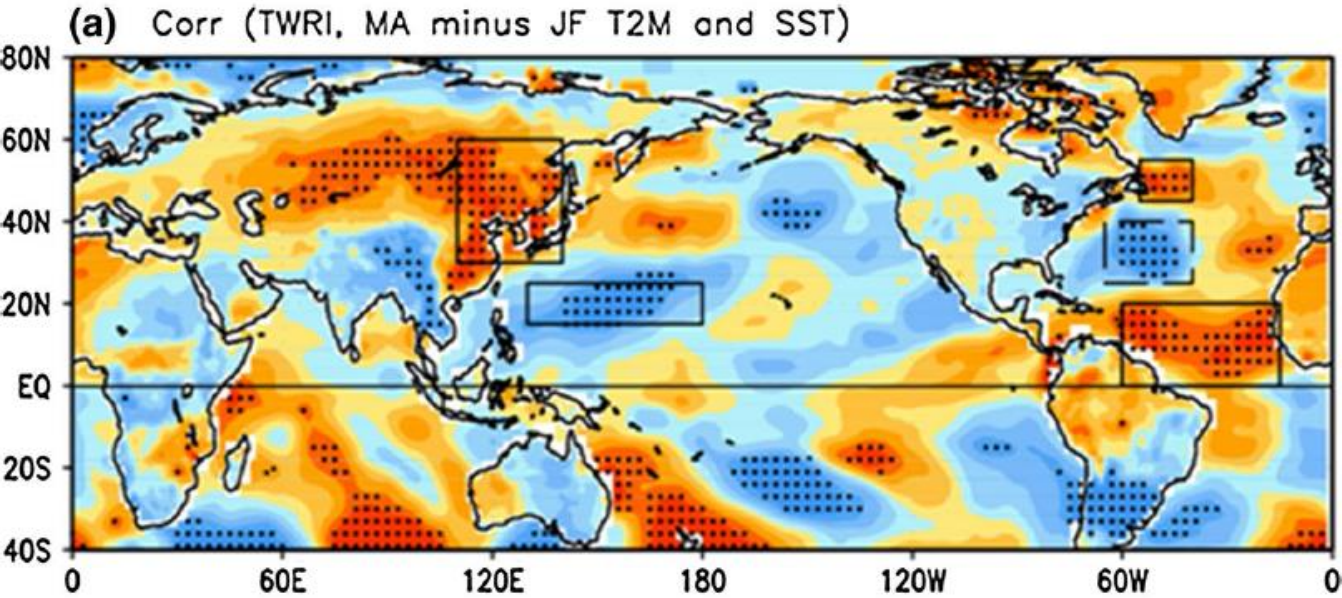
預報階段

模型輸入：
ERA5 重分析資料

統計預報模型

模型輸出：
臺灣梅雨季雨量預報

Meiyu Rainfall Forecast by P-E model



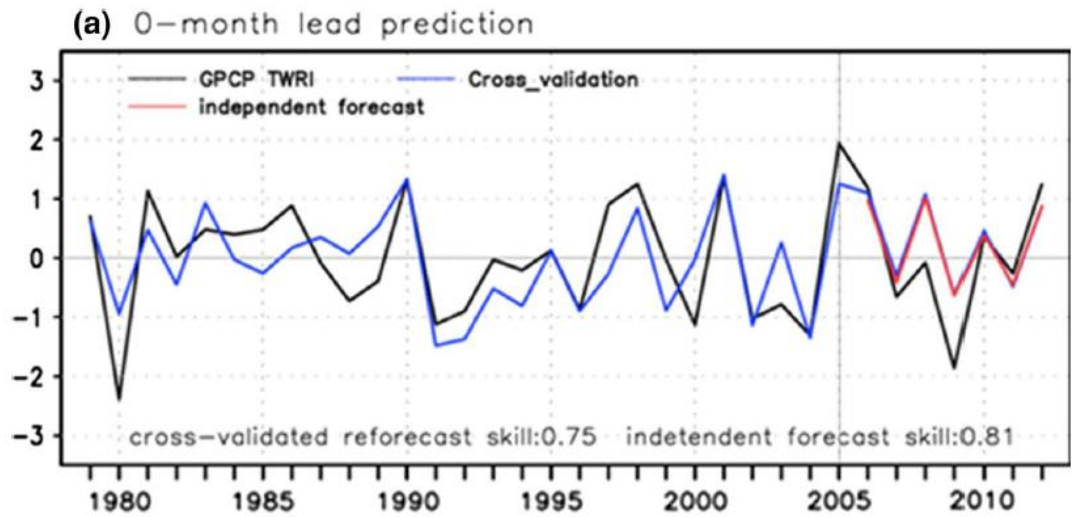
The 0-month lead forecast model identifies three key physical predictors for a strong Taiwan Meiyu based on March–April minus January–February tendency fields:

- 1. Cooling SST tendency over the western North Pacific
- 2. Mmeridional tripolar SST tendency across the North Atlantic
- 3. Warming tendency over northeastern East Asia.

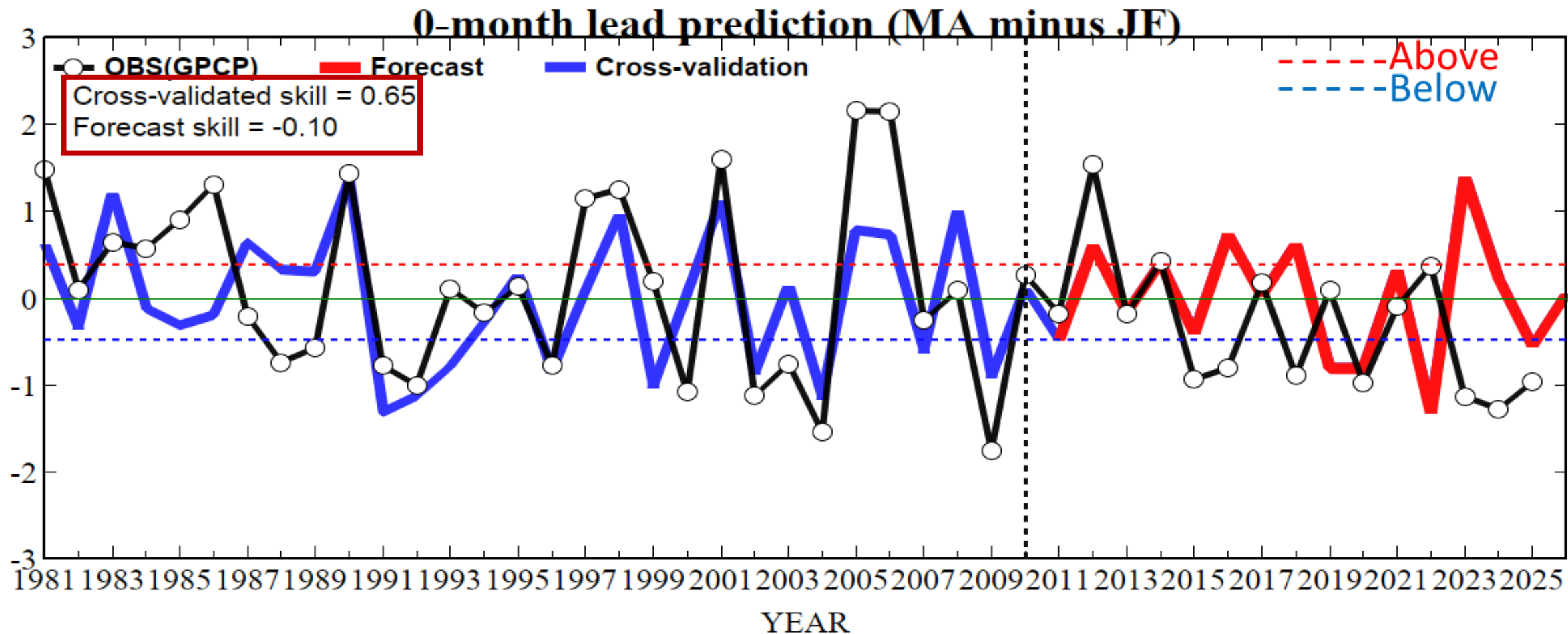
Table 2 The correlation coefficients (CCs) with the predictand (GPCP-TWRI) and among the predictors for the period 1979–2012

	TWRI	WNPT	NAT	EAT
TWRI	1.0	−0.50	0.47	0.46
WNPT	–	1.0	0.15	−0.15
NAT	–	–	1.0	0.17
EAT	–	–	–	1.0

The bolded numbers represent significance at 95 % confidence level



2026 MJ Meiyu Precipitation Forecast (GPCP)



臺灣梅雨季雨量在2026年預報為**正常**類別，但校驗2011-2025年的相關係數明顯偏低，主要原因是預報因子在**近十年**的相關係數值明顯偏低。

Predictors (Correlation Map)

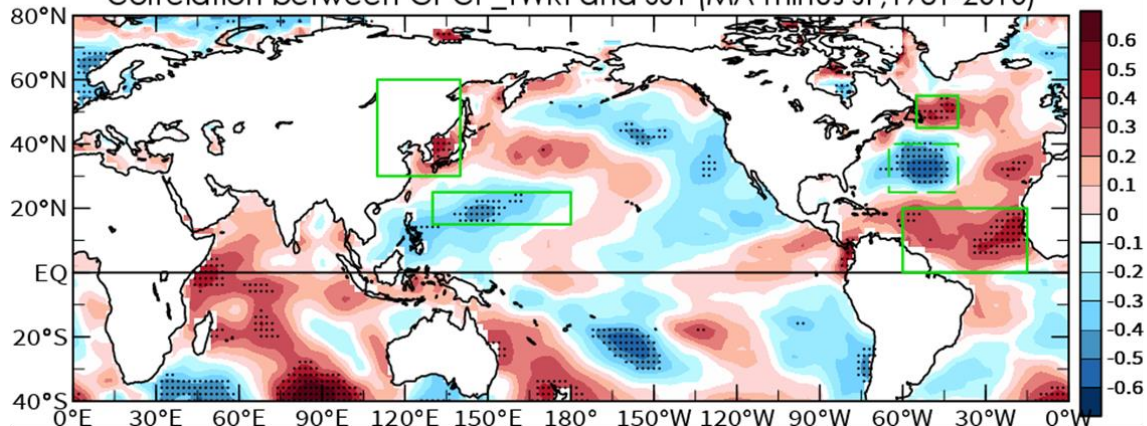


SST

Correlation Map between Index and SST(1981-2010)

Lead-0 Month

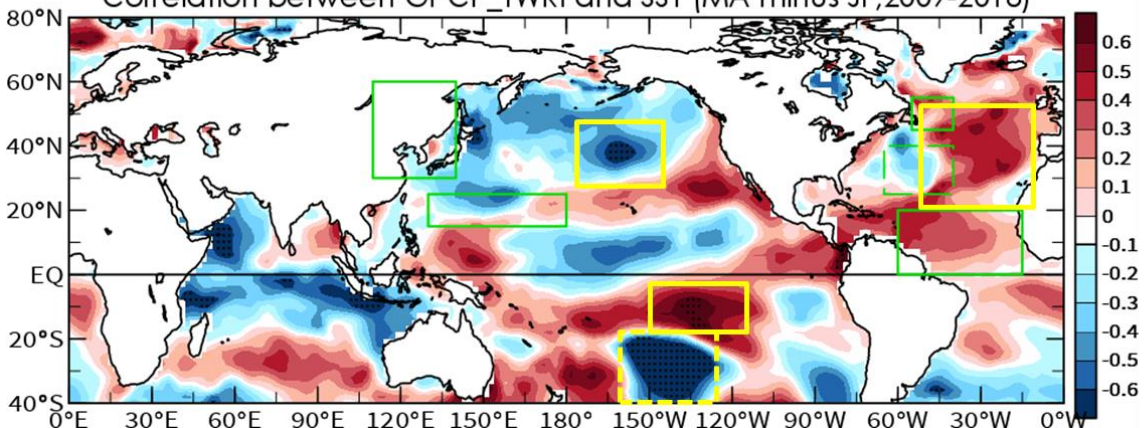
Correlation between GPCP_TWRI and SST (MA minus JF, 1981-2010)



Correlation Map between Index and SST(2009-2018)

Lead-0 Month

Correlation between GPCP_TWRI and SST (MA minus JF, 2009-2018)

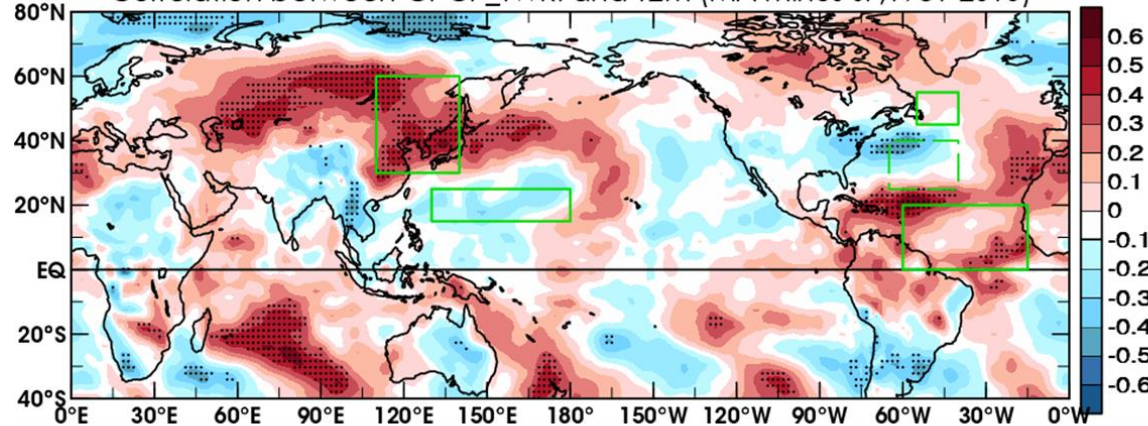


T2M

Correlation Map between Index and T2m(1981-2010)

Lead-0 Month

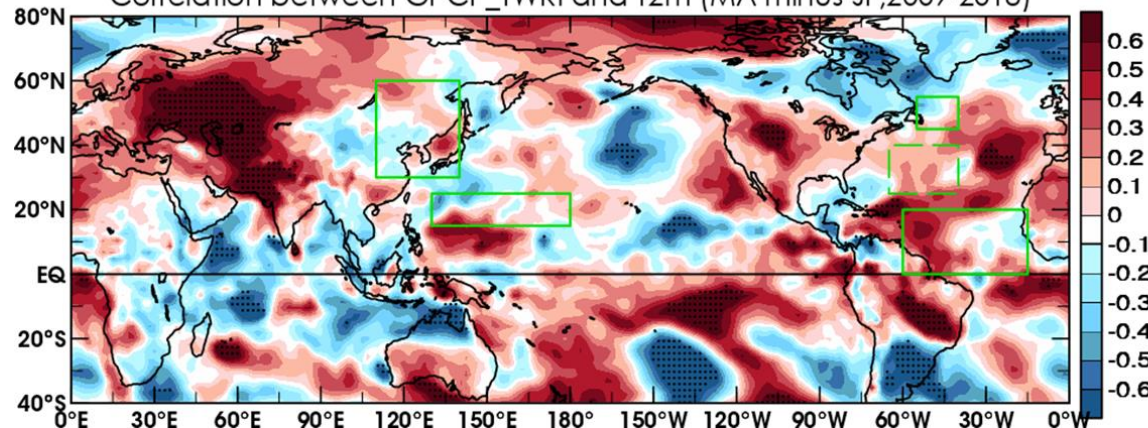
Correlation between GPCP_TWRI and T2m (MA minus JF, 1981-2010)



Correlation Map between Index and T2m(2009-2018)

Lead-0 Month

Correlation between GPCP_TWRI and T2m (MA minus JF, 2009-2018)



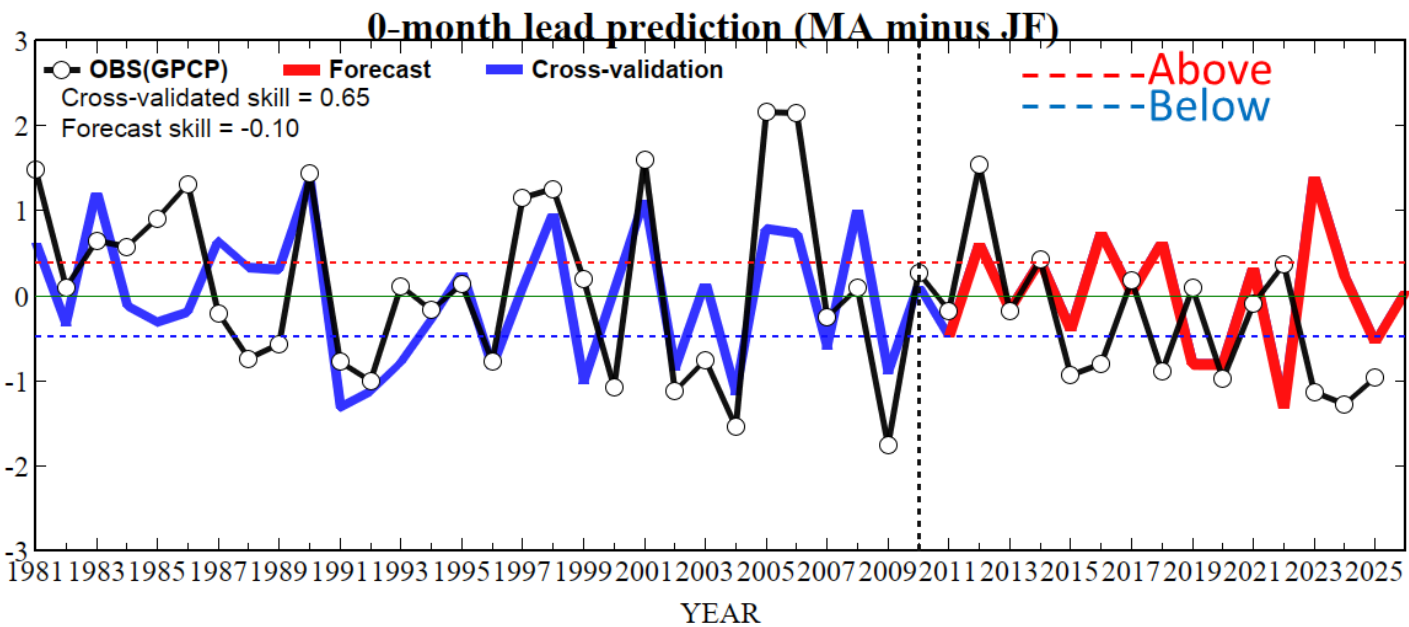
Choose SST region for New Predictors (yellow box)

2026 MJ Meiyu Precipitation Forecast (GPCP)

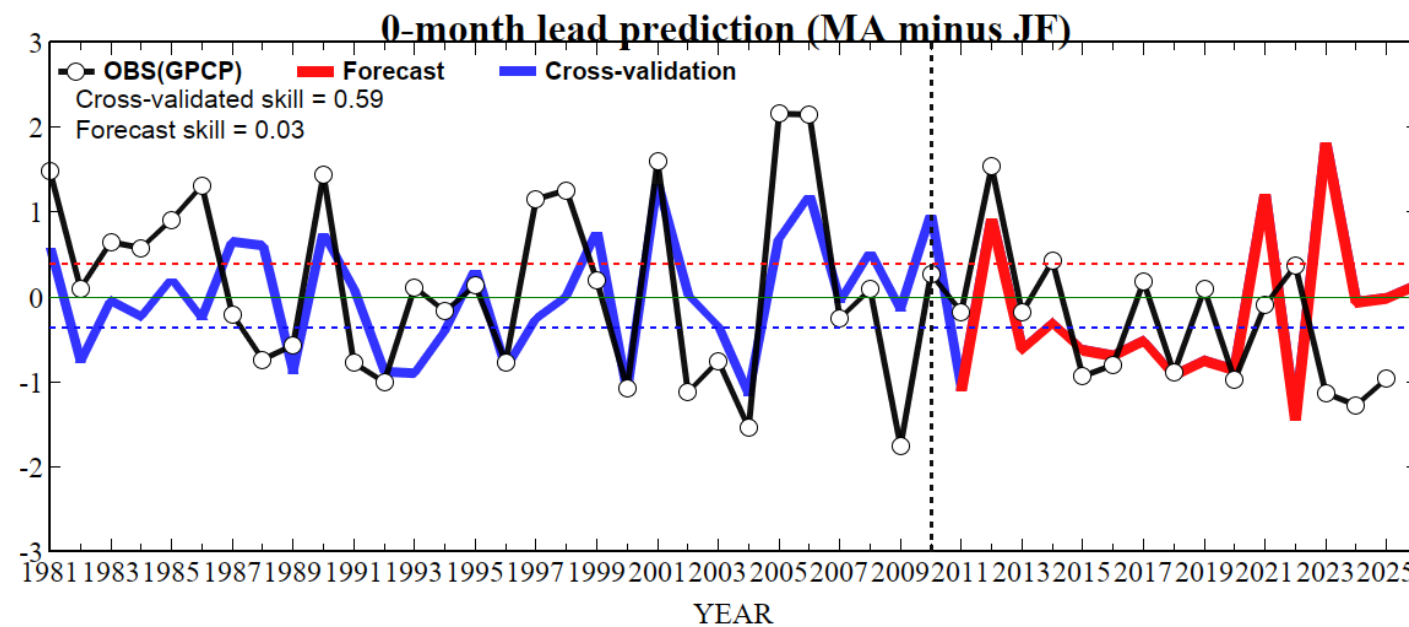


Lead-0 Month

Predictors



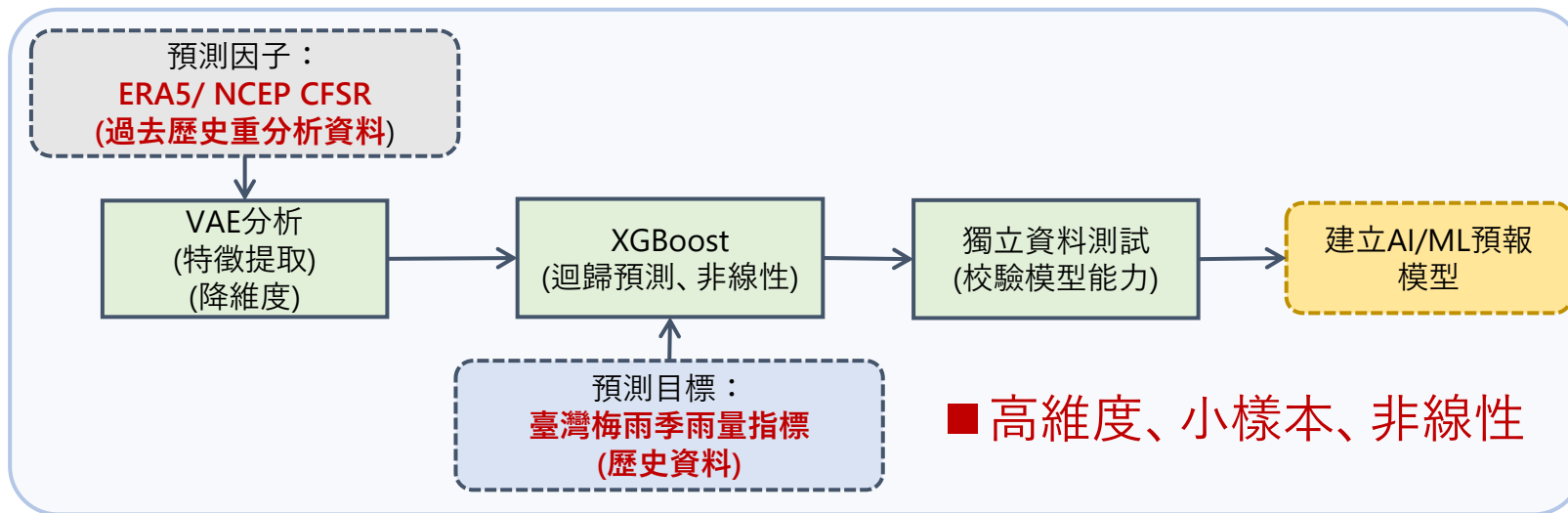
New Predictors



Meiyu Rainfall Forecast by AI/ML model



➤ AI/ML Model (VAE+XGBoost)



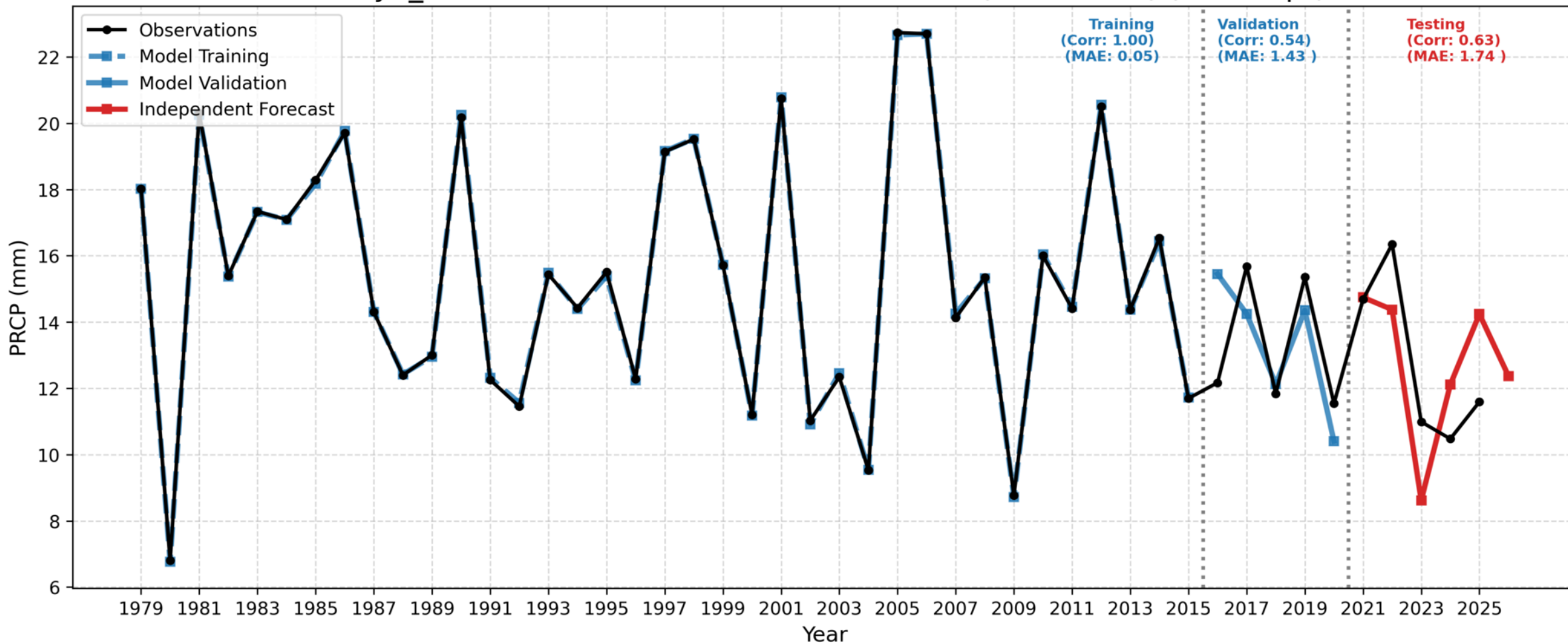
1. 資料: 1979-2025 ERA5 (SST, T2M, MSLP, SHUM850, Z200, U850, V850)
2. 預報對象: GPCP (area mean over 119E-123E, 21N-26N)
3. 訓練: 19798-2015; 校驗: 2016-2020; 測試: 2021-2025
4. 預報因子: 9個特徵
(VAE方法將7個氣象變數與範圍40E-120W, 20S-50N取出24特徵, 從24特徵選出與MJ梅雨季雨量有高相關的9個特徵)
5. 非線性迴歸預測模式: XGBoost



Meiyu Rainfall Forecast by AI/ML model

- 結合了深度學習（非線性特徵提取）與集成學習（強效迴歸預測）的高階預報架構。
- 將 VAE 與 XGBoost 結合，是為了解決氣象資料中「**高維度(3D)**、**小樣本(47年)**、**非線性**」這三大核心問題。
- 使用2026年1-4月的7個氣象變數透過AI方法預報2026年為正常-偏少類別

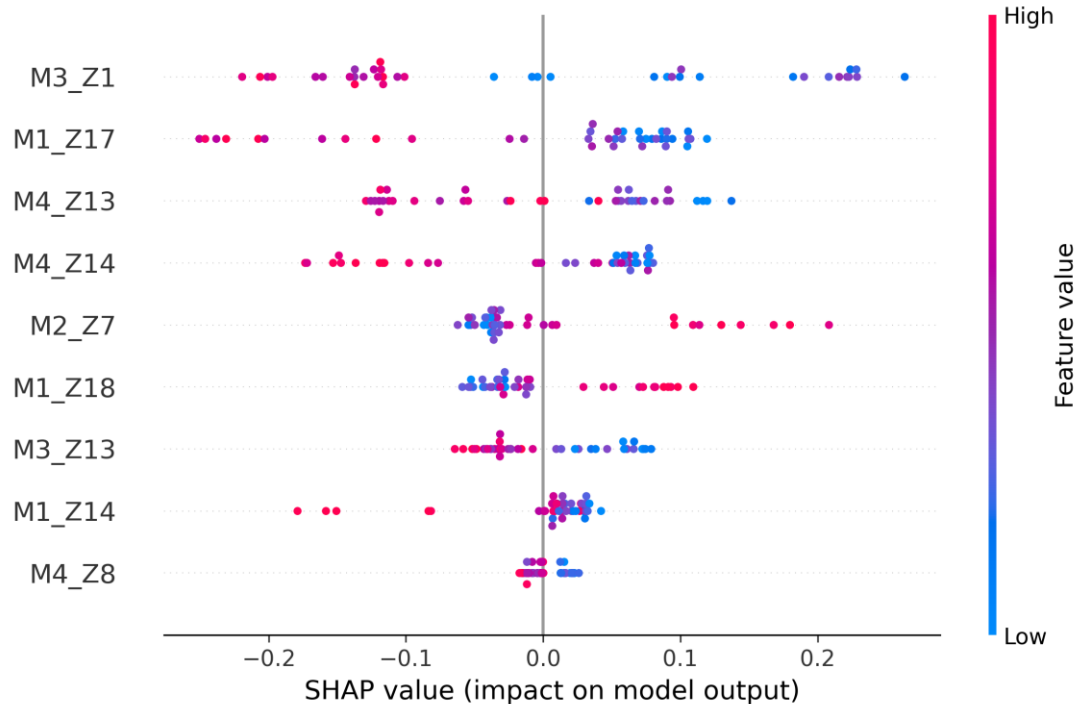
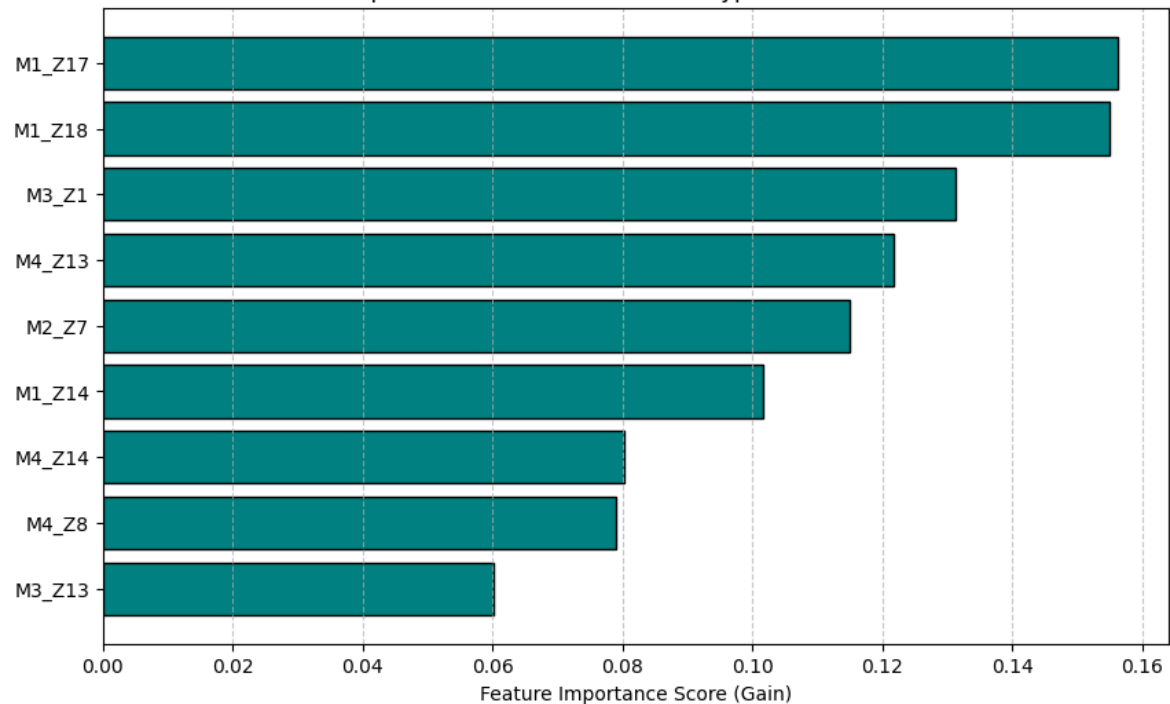
Meiyu_PRCP Index: Observation vs Prediction (1979-2026) (initial:apr)





特徵分析(shape套件)

Top 15 Crucial VAE Features for Typhoon Prediction



✓	M1_Z17	相關係數: -0.3425 XGB貢獻度: 0.1562
✓	M1_Z18	相關係數: +0.3117 XGB貢獻度: 0.1550
✓	M3_Z1	相關係數: -0.2506 XGB貢獻度: 0.1312
✓	M4_Z13	相關係數: -0.2792 XGB貢獻度: 0.1216
✓	M2_Z7	相關係數: +0.2764 XGB貢獻度: 0.1150
✓	M1_Z14	相關係數: -0.2647 XGB貢獻度: 0.1015
✓	M4_Z14	相關係數: -0.2660 XGB貢獻度: 0.0803
✓	M4_Z8	相關係數: -0.2729 XGB貢獻度: 0.0790
✓	M3_Z13	相關係數: -0.2668 XGB貢獻度: 0.0601



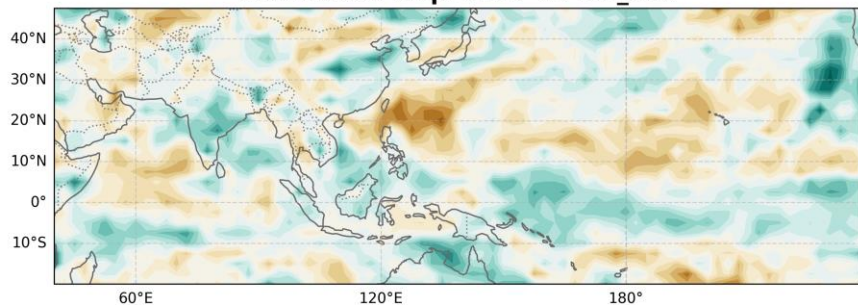
Correlation Map Between M1_Z17, M3_Z1 and MJ Variables



M1_Z17

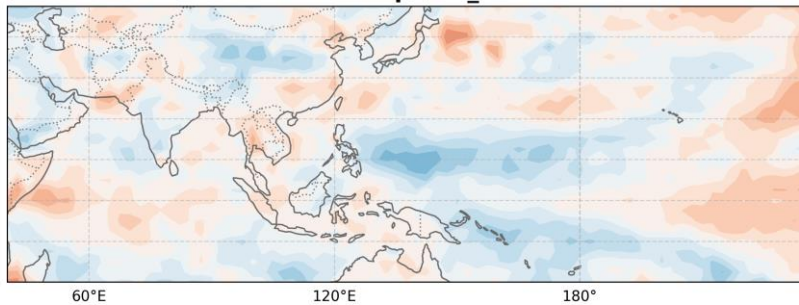
(**相關係數: -0.3425** | XGB貢獻度: 0.1562)

Correlation Map: PRCP vs M1_Z17



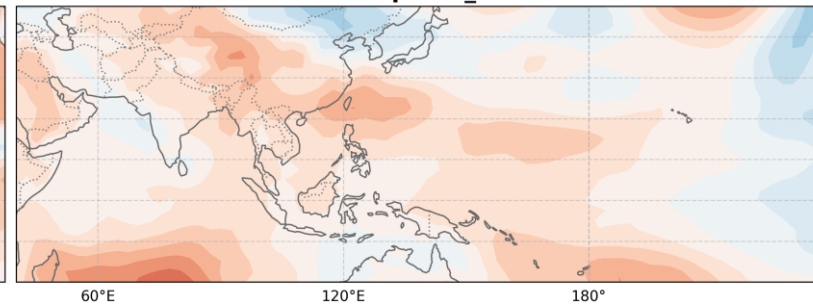
Pearson Correlation Coefficient (r)

Correlation Map: M1_Z17 vs SST



Pearson Correlation Coefficient (r)

Correlation Map: M1_Z17 vs MSLP



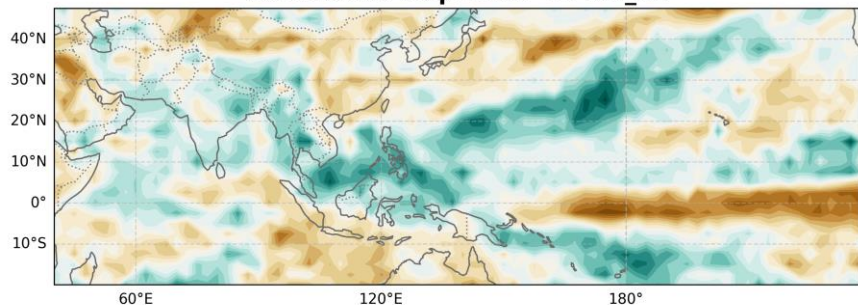
Pearson Correlation Coefficient (r)



M3_Z1

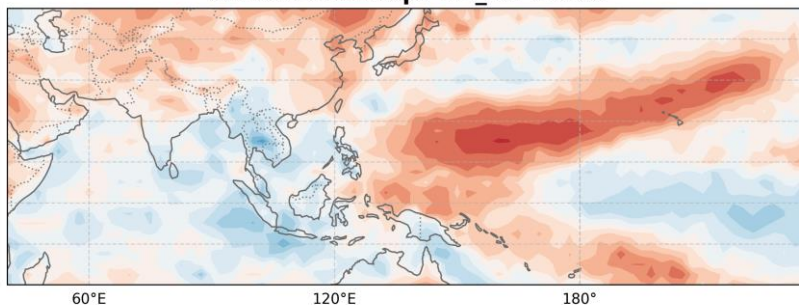
(**相關係數: -0.2506** | XGB貢獻度: 0.1312)

Correlation Map: PRCP vs M3_Z1



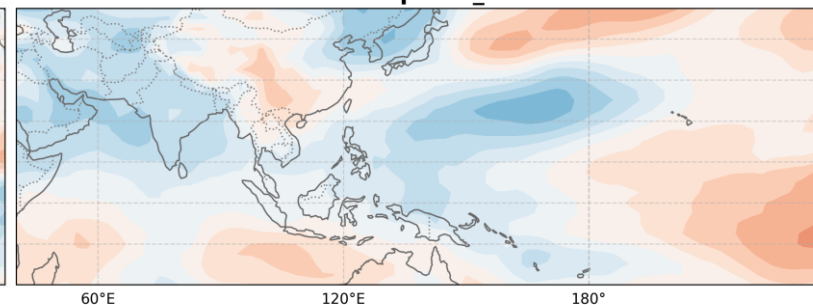
Pearson Correlation Coefficient (r)

Correlation Map: M3_Z1 vs SST



Pearson Correlation Coefficient (r)

Correlation Map: M3_Z1 vs MSLP



Pearson Correlation Coefficient (r)



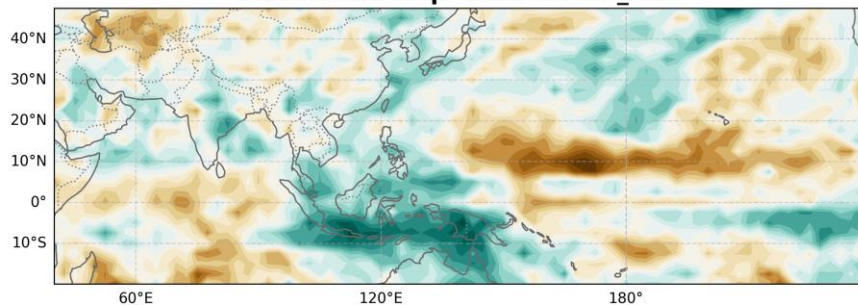
Correlation Map Between M1_Z18, M2_Z7 and MJ Variables



M1_Z18

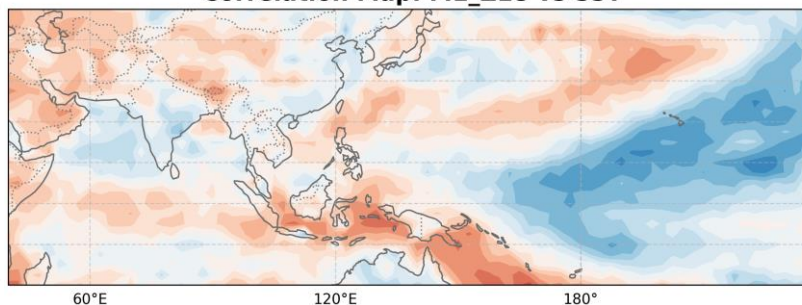
(**相關係數: +0.3117** | XGB貢獻度: 0.1550)

Correlation Map: PRCP vs M1_Z18



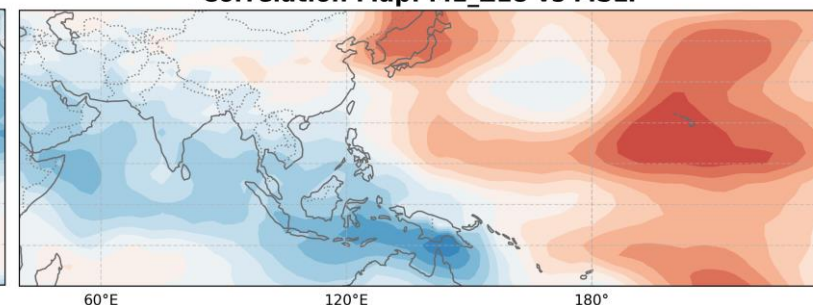
Pearson Correlation Coefficient (r)

Correlation Map: M1_Z18 vs SST



Pearson Correlation Coefficient (r)

Correlation Map: M1_Z18 vs MSLP



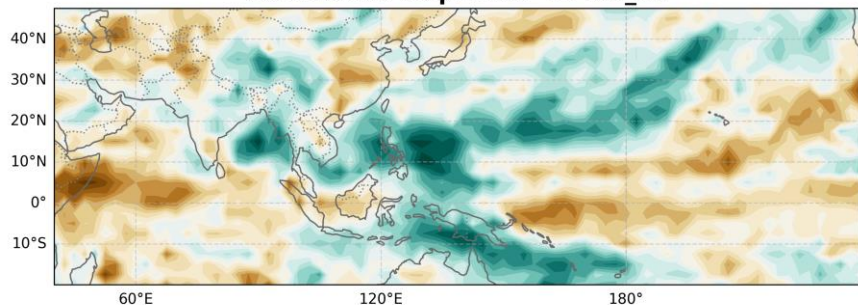
Pearson Correlation Coefficient (r)



M2_Z7

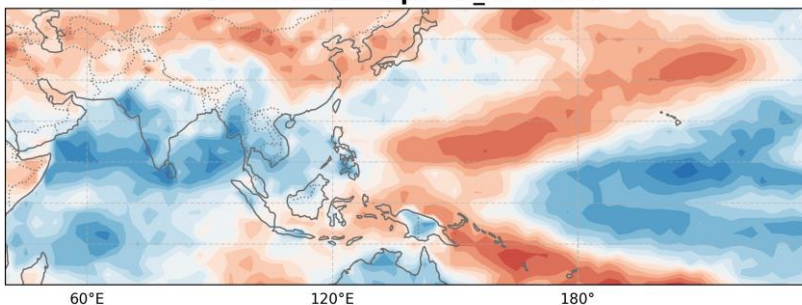
(**相關係數: +0.2764** | XGB貢獻度: 0.1150)

Correlation Map: PRCP vs M2_Z7



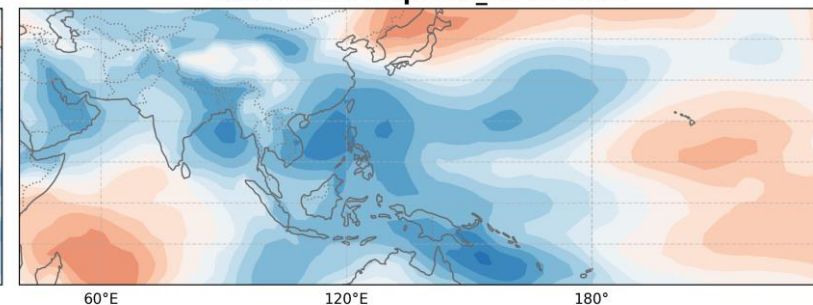
Pearson Correlation Coefficient (r)

Correlation Map: M2_Z7 vs SST



Pearson Correlation Coefficient (r)

Correlation Map: M2_Z7 vs MSLP



Pearson Correlation Coefficient (r)



- 1) PE-Model 由於訓練期與預報期間的預報因子相關性有明顯反相位，導致預報能力呈現負相關。
- 2) VAE可以有效降低維度及特徵提取，XGBoost有非線性作用與有效處理小樣本數據，透過此兩方法來建立的梅雨季雨量在訓練、校驗、測試期的相關係數值為0.96, 0.54與0.63 (MAE: 0.05, 1.43, 1.74)。
- 3) 未來將進行以下的測試
 - ✓ 加入海溫的差異值為預報因子(MA – JF sst)
 - ✓ 不同範圍大小的特徵提取之敏感度實驗
 - ✓ 利用氣候模式預報資料來進行AI建模(多模式、多系集、有限樣本數)

Thanks for your listening!

方法	EOF	MV-EOF	AE	VAE
處理關係	線性	線性	非線性	非線性
變數數量	單一變數為主	多變數 (需嚴格標準化)	多變數	多變數
解釋變異量	嚴格遞減排序 (可明確算出 %)	嚴格遞減排序 (基於標準化數值)	無排名機制 (整體重建)	無排名機制
特徵限制	強制正交 (互相垂直)	強制正交 (互相垂直)	無限制	無限制 (且特徵自動解譯)
潛在空間	特徵向量(PCA)	特徵向量(PCA)	破碎、不連續	平滑、連續分佈
生成解碼能力	無法憑空生成	無法憑空生成	極差 (易生出雜訊)	極佳 (能畫出合理物理場)
氣象應用極限	只抓大尺度主模態	抓多變數線性組合	僅作資料壓縮降維	挖掘非線性遙相關與極端天氣觸發機制
最大優點	運算極快、物理意義明確	能同時找尋多變數共變異	打破線性限制	生成能力強、完美對應氣候非線性演化
最大缺點	高模態常為無意義雜訊 解釋變異量很小	依然無法捕捉非線性機制	容易過度擬合、解碼常生出雜訊	訓練門檻高、需額外解釋特徵物理意義



- XGBoost (Extreme Gradient Boosting; 極限梯度提升器)
 - 有效處理小樣本數據: XGBoost 基於決策樹集成，在小樣本表現遠優於深度學習。
 - 特徵重要性: XGBoost 可明確說明某一個特徵對預報的貢獻度是多少，可以詳細分析特徵物理意義。
 - 非線性作用: XGBoost 會自動處理這些高度相關變數，不會產生數值不穩定的問題。

演算法	學習方式	非線性處理	運算速度	預測精準度	最佳適用情境
線性迴歸	畫一條直線	✗ 極差	⚡ 極快	低	尋找單純的總體趨勢
羅吉氏迴歸	S型機率曲線	✗ 極差	⚡ 極快	低	建立簡單的基準線 (Baseline)
決策樹	流程圖切割	☑ 佳	⚡ 極快	中低 (易過擬合)	需要產生簡單商業規則時
隨機森林	民主投票	☆ 極佳	🐢 慢	高	不想調參數、求穩定時
XGBoost	從錯誤中學習 (接力)	🏆 霸主級	⚡⚡ 超快	極高	追求極致預測準確度 找出掘複雜交互作用，需詳細調參數