

AI 輔助氣象預報——從技術革新邁入決策共生

賴曉薇^{1*} 王世宇² 洪景山³ 潘琦³ 陳孟詩³ 汪漢鈞¹ 蔡志賢¹

景丰科技股份有限公司¹

中央研究院環境變遷研究中心²

交通部中央氣象署³

(中華民國 114 年 10 月 21 日收稿；中華民國 115 年 5 月 6 日定稿)

摘要

2023 年 Google DeepMind 之 GraphCast 模型採用圖神經網路架構建構 AI 氣象預報模式 (Artificial Intelligence Weather Prediction, AIWP)，其預報準確度在 90% 預報變數上超越歐洲中期天氣預報中心之傳統數值模式，運算速度提升達千倍。本文針對近 3 年 AI 技術引發之預報結構變革進行系統性綜述研究。本研究依據時間尺度與作業角色建立 AIWP 分類框架，涵蓋即時預報 (0-6 小時)、中期預報 (1-10 天) 及次季節至季節 (Sub-seasonal to Seasonal, S2S) 預報等範疇，系統性回顧自端到端模型至混合架構之技術演進，並探討 Aurora 與 GenCast 等新的系統如何運用基礎模型克服傳統限制。在挑戰性較高的 S2S 領域中，FuXi-S2S 模型透過流場相關的擾動方案，將 Madden-Julian Oscillation, (MJO) 有效預報時效延長至 36 天。初代 AIWP 本質受限於均方誤差 (Mean Squared Error, MSE) 損失函數及缺乏物理守恆約束，普遍面臨預報場平滑化及極端強度系統性低估 (如熱帶氣旋強度低估 20-50%、極端降水偏差 35%)；因此，整合物理機制之混合架構已成技術演進的明確方向。針對臺灣氣象作業，本研究建議宜採取「以國際先進模型為基底、在地化校正技術為核心」策略，優先建置高品質臺灣區域再分析資料庫並發展高解析度降尺度技術，並拓展氣候風險治理。

關鍵字：人工智慧氣象預報、深度學習、氣候預報、極端天氣、降尺度技術

*通訊作者：賴曉薇 (hiaubi@simenvi.com.tw)

一、緒論

自古以來，人類對於天氣現象的預測始終懷抱深切期望。從古代觀雲識天之經驗法則，演進至當代以超級電腦進行數值模擬之科學方法，氣象預報本質上係一項對抗時間限制與運算能力極限之競賽。傳統建立於流體力學與熱力學之物理定律基礎上的數值天氣預報 (Numerical Weather Prediction, NWP)，需耗費龐大運算資源來模擬大氣網格點的時空演變 (Bauer et al., 2015)；此一預報作業模式歷經數十年發展，已成為全球氣象作業之主流方法。

然而，2023 年春季，AI 氣象預報 (Artificial Intelligence Weather Prediction, AIWP) 跳脫了傳統束縛，帶來革命性質變。NVIDIA 最先展示了 FourCastNet 系統 (Pathak et al., 2022)，隨後華為 (Huawei) 發表 Pangu-Weather 模型刊載於《Nature》期刊 (Bi et al., 2023)，Google DeepMind 研發的 GraphCast 模型也幾乎同時發表於《Science》期刊 (Lam et al., 2023)，這些研究成果從本質上改變了氣象預報的技術範疇。相較於 NWP 透過求解偏微分方程組以模擬大氣演變，AIWP 採用資料驅動 (data-driven) 方法，直接從歷史再分析資料中學習大氣系統之演變規律。實證結果顯示，此等 AI 模型之預報準確度已達到

甚至超越傳統頂尖預報系統的水準，同時運算速度提升達千倍：單一圖形處理器 (Graphics Processing Unit, GPU) 可於數分鐘內完成十天全球預報，相比之下，傳統方法需動用數千個處理器運算數小時才能達成相同任務 (Lang et al., 2024)。

此一技術革命迅速推進至作業化應用。ECMWF 於 2024 年正式將 AI 預報系統 (AI Integrated Forecasting System, AIFS) 納入日常作業流程，與其經典的綜合預報系統 (Integrated Forecasting System, IFS) 並行運作 (Lang et al., 2024)。至 2025 年，美國國家海洋暨大氣總署 (National Oceanic and Atmospheric Administration, NOAA) 正式啟用基於 GraphCast 架構的 AI 全球預報系統 (AI Global Forecast System, AIGFS；NWS, 2025)。在颱風路徑預報上，AIFS 的誤差只有 IFS 的 85% (ECMWF, 2025)，而 Beryl 跟 Helene 颶風的個案證明 GraphCast 具備實際作業的預報效能 (Sudharsan et al., 2025)，這些國家級單位的轉變，意味著 AI 的地位從「後處理輔助工具」跳躍至「獨立作業系統」之世代交替。

對臺灣而言，這樣的技術發展具有特殊之急迫性與重要性。臺灣地處西北太平洋颱風主要路徑之第一線，加以氣候變遷導致極端天氣事件頻率與強度持續增加，

氣象預報已從傳統之「天氣資訊服務」提升至「氣候風險治理」之應用層次。AIWP 的高效運算能力與前瞻預報特性，恰可提供災害發生前更長的預警前置時間。然而關鍵問題在於：臺灣氣象作業體系應如何有效整合此一新興技術？

有鑑於此，本研究選在技術變革之關鍵時刻，針對 AIWP 的科學演進進行系統性的回顧。本文之核心論述分為三個層面：首先，建立多維度之分類框架，依據預報時間尺度與作業角色定位，釐清各類 AI 模型之技術特徵與適用情境；其次，進行技術比較分析，系統性對比第一代代表性模型（如 GraphCast、Pangu-Weather、AIFS）與次世代生成式系統（如 Aurora、GenCast）之架構差異與物理假設，並探討 FuXi-S2S 模型如何突破次季節預報之傳統限制。最後，誠實評估 AIWP 在平滑化與物理一致性上之方法論侷限，並結合臺灣特殊之複雜地形與氣候治理需求，提出切實可行之應對策略建議。本文為一篇回顧性研究，期能協助關注此領域之研究人員、作業預報員及政策決策者於技術浪潮中找到定位，並為臺灣氣象預報作業的轉型提供方向。

二、AI 氣象預報模型之系統性分類

當 GraphCast、Pangu-Weather 等突破性模型相繼問世時，學術界與作業界對於此類技術在預報體系中之確切定位尚缺乏完整理解。為協助實務工作者（特別是作業預報員）釐清面對特定預報需求時應選擇何種工具，本章建立一套分類框架，藉由三個角度構築立體座標系統，以定位各 AI 模型之生態位（niche）：第一個視角在於審視 AIWP 如何在傳統預報核心區域實現性能提升；第二視角分析次世代基礎模型如何突破「可預報度間隙」並延伸至次季節至季節（Sub-seasonal to Seasonal, S2S）領域；第三視角則關注技術整合的路徑，探討差異化整合與不同尺度的銜接。

（一）預報效能之時空擴張

即時預報（nowcasting，0-6 小時）是 AI 氣象應用中技術發展中最早也最為成熟之領域。相較於傳統 NWP 受限於複雜之資料同化（data assimilation）與模式初始啟動（spin-up）過程，AI 模型繞過物理方程的求解步驟，直接解析雷達回波與衛星雲圖，實現近乎即時之天氣系統推演。在技術架構方面，早期主流方法多採用卷積神經網路（Convolutional Neural Networks, CNN）進行模式識別（Sønderby et al., 2020），其後引入生成對抗網路（Generative Adversarial Networks, GAN）則有效解決預

報模糊化問題 (Ravuri et al., 2021)。這類模型是透過學習海量歷史案例建立預測映射關係，因此當遭遇訓練資料集中罕見之非典型極端天氣事件時，模型往往因缺乏物理機制的顯式約束而導致預報失準，此一現象凸顯未來融合物理知識 (physics-informed) 之重要性。

中期預報 (1-10 天) 是氣象預報最核心、社會需求最為強烈之時間範圍，也是近年 AI 技術突破之重心。當純資料驅動的 AI 模型宣稱能與氣象界的黃金標準——ECMWF 之預報系統分庭抗禮甚至超越時，對學術界與作業界造成之衝擊不言而喻。其中，GraphCast 展現一種革命性創新，大膽摒棄傳統 NWP 沿用之經緯度網格系統，改採二十面體多層次結構 (icosahedral multi-resolution mesh) 切割地球表面，利用圖神經網路 (GNN) 優異之節點連結能力捕捉大氣長程交互作用 (long-range interactions)。GraphCast 基於 1979-2017 的 ERA5 再分析資料 (ECMWF Reanalysis v5) 訓練的結果顯示：在 1,380 個驗證指標中，有超過 90% 顯著優於 ECMWF 之高解析度預報系統 (High Resolution forecast, HRES) (Lam et al., 2023)。然而，隨預報時效延長，GraphCast 預報場出現模糊化現象；另一方面，在平流層之預報技巧也相對較低，此一限制源於訓練資料在平流層

的資料密度較低，以及模型架構對垂直結構之解析能力有限。

Pangu-Weather 則採用三維地球專用 Transformer 架構 (3D Earth-specific Transformer)，將不同壓力層的大氣狀態整合為三維立方體結構，得以兼顧水平與垂直方向變化。藉由層級時間聚合策略 (hierarchical temporal aggregation) 訓練多個專門化模型，並採取分時合作串聯不同預報延時 (1 小時、3 小時、6 小時及 24 小時；Bi et al., 2023)。Pangu-Weather 在西北太平洋熱帶氣旋預報表現優異，路徑預報改善幅度達 8-19%；然而強度預報則呈現 15-20% 之系統性低估。

相對於科技企業之革命式創新，ECMWF 所開發之 AIFS 代表謹慎與融合之作業化典範。ECMWF 從設計初始即將 AIFS 定位為 IFS 之速度夥伴與系集生成器，而非競爭對手 (Lang et al., 2024)。AIFS 採用 GNN 與 Transformer 混合架構，直接於 IFS 原生網格運作，透過三階段訓練策略確保模型學習大氣運動基本規律並適應當前作業系統需求。也因此，其作業優勢展現於時間效益與系集預報能力：AIFS 能於 IFS 完成資料同化前提供初步預報，節省約 90 分鐘；並可於數分鐘內產生 50 個系集成員，有助於預報員迅速掌握不確定性分布 (Mahesh et al. 2025)。

FourCastNet(FCN)在工程設計上則是極致優化其運算效率。該模型採用自適應傅立葉神經算子(Adaptive Fourier Neural Operator, AFNO)，其核心創新在於融合偏微分方程求解器的數學原理與深度學習架構，透過快速傅立葉變換將運算複雜度從 $O(N^2)$ 降至 $O(N \log N)$ ，實現與圖形處理器並行運算架構之高度適配。性能測試顯示，FCN可於2秒內完成一週全球預報，運算速度較傳統NWP快約4.5萬倍，能源消耗僅為IFS之1/12000(Pathak et al., 2022)。然而，FCN在極端降水事件預報上平均低估35%，低估幅度較其他模型更為顯著，此一現象可能與其過度追求運算效率而在模型容量與表達能力方面有所妥協相關。

(二) 次世代模型

AI氣象預報技術分化出兩條截然不同之演進路線。Microsoft推出的Aurora可作為「基礎模型」(Foundation Model)代表範例。Aurora於2025年5月發表於Nature並開源化(Bodnar et al., 2025)，借鑑大語言模型之成功策略，先於廣泛之地球系統資料集(大氣、海洋、冰層)進行大規模預訓練以學習通用表徵，再針對特定任務進行微調。如此可有效利用大氣、海洋與冰層間的物理共通性，實現跨領域知識遷移(Bodnar et al., 2025)。它的技術創新顯示在多模態整合能力——可同時處

理不同解析度與不同物理領域之資料，展現靈活適應性，得以快速適配至區域高解析度預報或特定極端事件預測的任務，即使在資料稀缺領域(如海冰預測)，也能透過遷移學習達到合理預報技巧。ECMWF已將其預報結果納入作業圖表目錄作為外部ML模型之一。

GenCast則選擇在機率生成領域尋求突破。該模型採用擴散模型架構，結合聯合學習策略與邊際化技巧，生成物理一致且統計校準良好之完整機率分布(Price et al., 2024)。GenCast的技術特點包括生成式建模產生多樣化且物理合理的預報場景、提供經過校準的機率資訊，有效改善極端值預報的系統性低估，在極端降水等高難度預報任務上展現出顯著優於迴歸模型之優勢(Price et al., 2025)。

在兩週至數月(次季節至季節尺度)時間尺度上，大氣系統對初始條件之記憶逐漸喪失，同時尚未完全受控於海洋等邊界強迫之主導；其可預報度主要來自海表面溫度(SST)異常等緩慢變異的邊界強迫，以及季內振盪(Madden-Julian Oscillation, MJO)所激發並調控的熱帶-中緯度遙相關。在這個預報時間域，既無法單純依賴初始值，亦難以憑藉氣候場推演，NWP的預報技巧呈現急劇下降趨勢(Vitart and Robertson, 2018)。這片「可預報度沙漠」

長期被視為最棘手的預報領域，隨著 AI 技術的引入，此領域出現了重要突破。FuXi-S2S 的創新方案主要包括兩大關鍵設計：首先，它充分利用初始時刻的 SST 異常作為邊界強迫，有效延長 MJO 的預報技巧；其次，它在模型的隱藏特徵（latent features）中引入流場相關擾動（flow-dependent perturbation scheme），大幅提升系集預報對大氣不確定性的描述能力（Li et al., 2024）。不同於傳統隨機擾動方法，FuXi-S2S 透過深度學習習得與當前大氣流場型態高度相關的擾動結構，從而更真實地反映不確定性的動力演化特徵。這項進步在 MJO 預報中尤為突出：將有效預報時效從傳統模式之 20-30 天大幅延伸至 36 天，北半球冬季更可達 42 天。以 2022 年巴基斯坦洪水為例，針對提前四週之初始化預報，傳統 ECMWF S2S 模式顯著低估降雨強度，僅達觀測值三分之一；相較之下，FuXi-S2S 預測之季節降雨異常值達 +3.8 個標準差，與觀測值（+4.0 個標準差）高度吻合，且空間分布相關係數更高達 0.82，這項成就大大提升早期預警的參考度。

(三) 多層次的整合路徑

在釐清預報時效之定位後，模型架構的「整合維度」成為界定 AIWP 發展路徑的另一關鍵。儘管純資料驅動之端到端模型具備極致之推理效率與部署優勢，但在

缺乏物理動力核心約束的情況下，單一時間步之微小誤差易透過自回歸過程不斷累積放大（Arcomano et al., 2020；Kochkov et al., 2024）。為尋求技術效能與物理合理性之平衡，AI 與物理機制之「混合整合（Hybrid）」成為技術演進之必然；依據整合的範疇，可分為作業並行、架構耦合與尺度銜接三方面。

ECMWF 之 AIFS 代表一種雙軌並行之系統整合典範。如前述，AIFS 僅需 5-10 分鐘即可產出結果，在時間效益上不但為預報員爭取約 90 分鐘之黃金預警時間，還能快速生成數百個系集成員提供機率預報，有效考慮極端天氣之可能性。AIFS 及 IFS 分別扮演「效率夥伴」與「物理基準」之角色，更關鍵的是，當兩套系統預報路徑出現分歧時，即可觸發人工審查機制，將預報差異轉化為具體決策價值（Lang et al., 2024）。NeuralGCM 則是在模型架構層面更徹底的混合策略。NeuralGCM 的動力核心保留傳統數值方法以確保流體力學之基本守恆律，而需要高計算成本之物理參數化過程則由 AI 學習替代，整個系統實現端到端之可微分訓練（end-to-end differentiable training），既利用 AI 之高效率與高準確度，又透過物理骨架解決純 AI 模型之長時效預報漂移問題（Kochkov et al., 2024）。

本文提倡的整合路徑也包含空間尺度層面—從全球 AIWP 本身進一步延伸到其生態系中的高解析度降尺度技術。降尺度技術在氣象領域早已行之有年，傳統上常見者為統計降尺度與動力降尺度，在 AI 技術引入後，近年來已發展出結合確定性與生成式模型的新型混合降尺度方法，在處理複雜地形與小尺度細節上展現顯著優勢。以臺灣地勢起伏劇烈之中央山脈案例而言，在 25 公里解析度之全球模式僅能由少數網格點粗略代表，致使關鍵之地形強迫與局地環流特徵無法被有效捕捉，更突顯了高解析度降尺度技術對臺灣提供精緻預報的重要性。中央氣象署與 NVIDIA 合作開發之 CorrDiff 模型展示生成式 AI 在此領域進展顯著。CorrDiff 模型採用兩階段設計，其核心概念為「殘差修正」

(Residual Corrective) 策略：第一階段利用 UNet 架構建立粗、細解析度場之間的基礎映射，第二階段則引入擴散模型以生成符合物理統計特徵之小尺度細節 (Mardani et al., 2025)。這種確定性映射及生成式修正的結合，讓模型能在保留大尺度背景場的同時，補充真實且多樣的小尺度變異。2023 年海葵颱風案例提供有力驗證：經 CorrDiff 降尺度後，中心最大風速之誤差改善幅度達 50%，強風半徑與實

際觀測高度吻合，並成功重建清晰之颱風眼牆結構。

三、作業可行性與個案分析

技術之先進性最終必須在作業現場接受實戰檢驗，蓋模型之價值乃取決於其在第一線之實際表現；即便一個模型在學術論文中展現優異之平均統計分數，也不意味著在現實預報環境中即可被無條件信任。所謂「作業可行性」(operational feasibility) 實為一項多維度之綜合考驗，涵蓋預報之持續穩定性、面對異常情境之應變能力、與現有工作流程之整合程度，以及其中最關鍵之要素——預報員之信任，即決策者是否敢於依據模型輸出採取實際行動。

為跨越學術研究與預報作業間的鴻溝，ECMWF 採取嚴謹之三階段部署策略，逐步累積系統之實戰信用。首階段 (2023 年 10 月至 2024 年 3 月) 為長達半年之「平行測試期」，期間 AIFS 與 IFS 同步運作但不對外發布，由預報員逐日比對兩者差異，仔細檢視 AIFS 在極端天氣事件中之穩定性，累積內部信心。其後進入第二階段 (2024 年 4 月至 6 月) 之「有限發布期」，向少數國家氣象機構提供標註為「實驗性」的產品，收集關於系統在不同應用場景下的可用性與可信度回饋。最終於 2024 年 7

月，AIFS 正式被納入例行作業參考指引，確立了其角色定位。此時，AIWP 與 NWP 形成完善的策略分工：AIFS 憑藉其運算效率優勢，專注於提供極速之初步態勢研判，而 IFS 則繼續發揮其物理精確性之強項，形成雙軌互補之預報體系（Lang et al., 2024）。2025 年 7 月，AIFS ENS 正式投入作業運行（ECMWF, 2025），以高速系集生成能力，在數分鐘內提供豐富的不確定性預報，顯著提升了對極端天氣事件發生機率與可能性空間的描述能力。透過早期警示、差異比對審查，以及系集資訊，兩套系統共同提升了預報的時效性、準確度與決策價值（Lang et al., 2024）。

美國 NOAA 的經驗則在凸顯「在地化訓練」（localized training）之重要性。此處所謂的「在地化」，並非指地理區域的局部訓練，而是指機構在地化（agency-specific）或作業系統在地化——讓 AI 模型適應 NOAA 自身的資料特性與作業流程。NOAA 於 2025 年 12 月發布服務變更通知（Service Change Notice 25-89），正式將 AIGFS v1.0 納入常規作業（NWS, 2025）。為克服預報員與公眾對「AI 黑盒子」之疑慮，NOAA 採取嚴謹之信任建構策略。AIGFS 先是捨棄 Google 基於 ERA5 訓練的原廠權重，轉而使用 NOAA 自有的全球資料同化系統（Global Data Assimilation

System, GDAS）資料進行重新訓練；如此不僅解決異質資料源導致之初始化不相容問題，更確保 AI 模型與現有物理模型在初始場的統計特性上之一致性，從而減少系統性偏差（Tabas, 2024）。在程序上，系統經歷長達一年之「實驗性平行運轉」，期間透過公開資訊聲明（Public Information Statement, PNS）廣泛徵詢意見，並定期發布與傳統 GFS 之並列驗證報告（NWS, 2024）。如此謹慎且透明之部署過程，顯示作業化系統對可靠性與可追溯性之高度要求。

經過嚴謹作業化流程驗證之系統，在面對高影響天氣系統時如何展現其作業價值與物理侷限？熱帶氣旋預報正是檢驗其核心能力之最佳試金石，因這類的預報同時考驗模型對大尺度環流引導與中小尺度對流發展的模擬能力。在路徑預報方面，AIFS 之 120 小時預報平均位置誤差約 250 km，優於傳統 IFS 之 300 km 以上，主要歸功於其修正傳統模式常見之移速過慢偏差。跨模式研究亦支持此趨勢，Sudharsan et al.（2025）指出，即便使用尚未經在地化微調之 GraphCast 模型搭配 GFS 初始場，在 Beryl 與 Helene 颶風個案中，其 5 天路徑誤差相較於 HWRFx 物理模型仍減少約 30-40%。此外，AIFS 在中長期極端降雨的預報（如 2025 年澳洲昆士蘭熱帶低

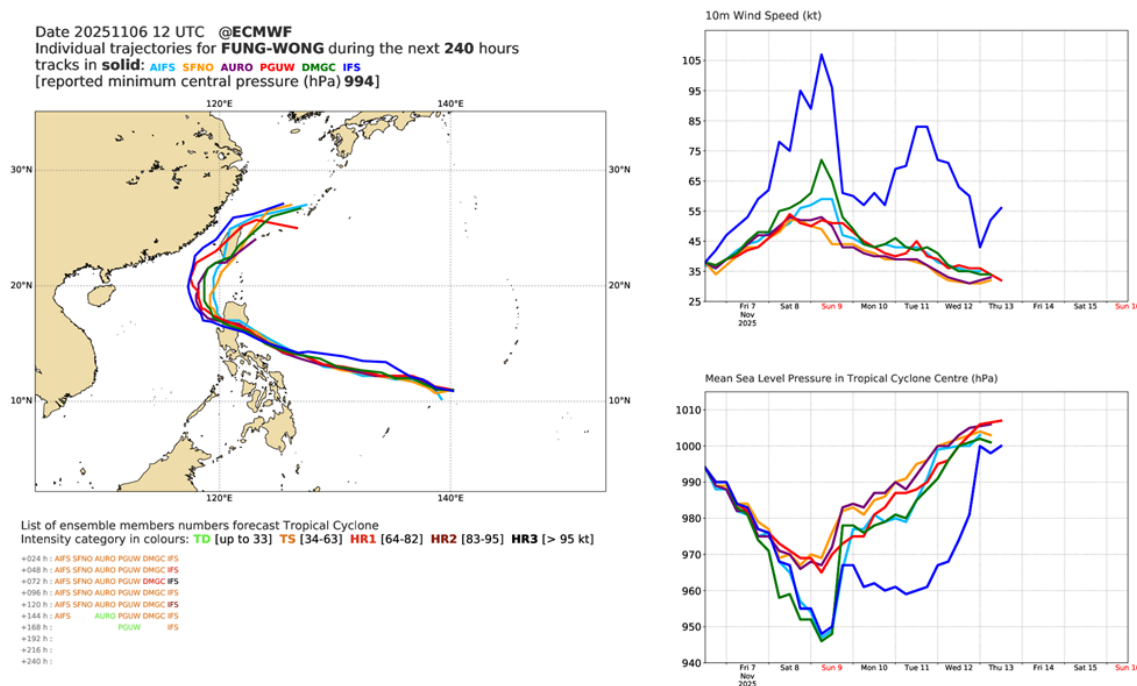


圖 1：2025 年 11 月 6 日 12 UTC 為初始場之鳳凰（FUNG-WONG）颱風預報路徑、10 米風速及中心氣壓，各 AI 模型與傳統 IFS（藍線）之比較。其中包括 ECMWF 自行開發的 AIFS（淺藍線），NVIDIA 的 FourCastNet（SFNO，橘線），Microsoft Aurora（AURO，紫線），華為雲的 Pangu-Weather（PGUW，紅線），Google DeepMind 的 GraphCast（DMGC，綠線）。圖片來源：取自 ECMWF Tropical Cyclone Tracks - Machine Learning，https://charts.ecmwf.int/products/cyclone?base_time=202511060000&product=tc_tracks_ml_forecast&unique_id=32W_FUNG-WONG_2025。

壓，提前 5 天預報出超過 300 mm 的雨量）展現捕捉中長期極端災害天氣之潛力。然而，相較於路徑預報之優異表現，統計資料則揭示 AI 模型在強度與結構細節上仍存在明顯之物理侷限。AIFS 在第 5 天之中心氣壓預報平均偏差約 10 hPa，高於 IFS 之 4 hPa，顯示其預報強度普遍偏弱（Lang et al., 2024, Figure 9）。

近期颱風案例可進一步檢驗新一代 AIWP 效能。以 2025 年 11 月菲律賓東方海面之鳳凰（FUNG-WONG）颱風為例，

AIFS ENS 在颱風生成前 48 小時的預報即有過半系集成員反映出擾動訊號。ECMWF 官網顯示 AIFS、GraphCast 與傳統 IFS 均能掌握颱風通過呂宋島後之大角度北轉趨勢，證實 AI 處理大尺度駛流場之能力已趨成熟（圖 1）。但此案例呈現出結構模擬之矛盾：AI 模型雖成功預報氣壓下降（至 955 hPa），但對應之風速僅達輕度颱風上限，遠低於聯合颱風警報中心（Joint Typhoon Warning Center, JTWC）觀測之 $60 \text{ m}\cdot\text{s}^{-1}$ 。此類「氣壓-風速脫鉤」現

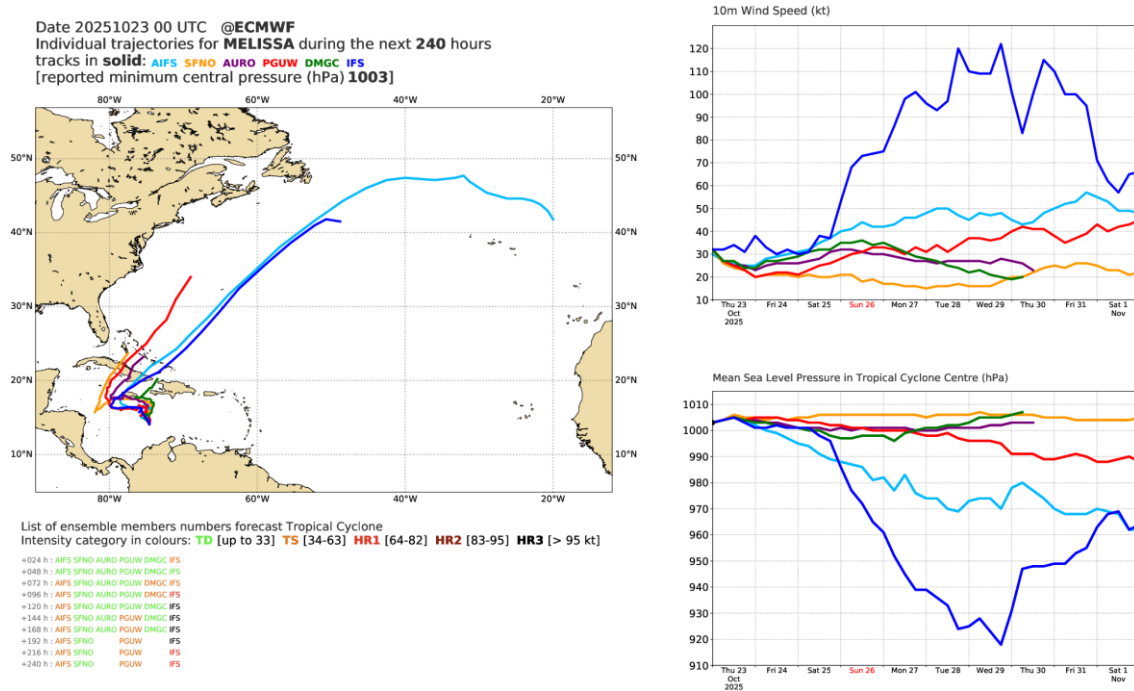


圖 2：同圖 1，但為 2025 年 10 月 23 日 00 UTC 為初始場之 Melissa 颶風預報比較。（圖片

來源：取自 ECMWF Tropical Cyclone Tracks - Machine Learning，

https://charts.ecmwf.int/products/cyclone?base_time=202510220000&prod-

[uct=tc_tracks_ml_forecast&unique_id=13L_MELISSA_2025](https://charts.ecmwf.int/products/cyclone?base_time=202510220000&prod-uct=tc_tracks_ml_forecast&unique_id=13L_MELISSA_2025)。

象，顯示目前 AI 模型在缺乏物理約束下，尚難以正確建構極端風場之梯度。2025 年 10 月 Melissa 颶風預報則面臨更加嚴峻的考驗，此創紀錄之五級颶風（中心氣壓 892 hPa、最大風速 $82 \text{ m}\cdot\text{s}^{-1}$ ）擊穿所有預報模式之天花板：面對 48 小時內風速從 $33 \text{ m}\cdot\text{s}^{-1}$ 躍升至 $58 \text{ m}\cdot\text{s}^{-1}$ 之快速增強（Rapid Intensification, RI）過程，幾乎所有模型都慘遭滑鐵盧。傳統 IFS 雖然能預警其快速增強趨勢，並出預報接近四級颶風之風速（ $55 \text{ m}\cdot\text{s}^{-1}$ ），但仍大幅低估巔峰強度；相比之下，AI 模型表現更為嚴重低估，包括 AIFS 與 GraphCast 在內之主流 AI 模式，在關鍵

期之預測峰值僅約 $28 \text{ m}\cdot\text{s}^{-1}$ ，幾乎完全漏報此場歷史級風暴（圖 2）。

AI 模型在強度預報上之困難主要受限於其訓練方式與物理機制之缺失。首先，以均方誤差為損失函數傾向於鼓勵模型預測接近氣候平均值，而習得較為保守之平滑策略（相關討論詳見第 4 節）。其次是訓練資料之解析度限制——以 ERA5 再分析資料為例，其 25 公里之網格間距無法解析熱帶氣旋之內核結構，而快速增強之關鍵物理過程往往發生在 10 公里以下之尺度。再者，快速增強過程涉及高度非線性之物理機制，包括眼牆對流之爆發性發展與海洋熱含量之快速釋放；在缺乏物理約束情

況下，純由資料驅動的模型往往難以捕捉此類變化過程 (Brenowitz et al., 2025)，而系集系統雖然同樣無法徹底解決低估強度的問題，但透過尾部分佈採樣，少數成員仍有機會捕捉到極端事件，提供具統計意義的早期預警訊號。此外，由於目前 AI 模型尚缺乏海氣耦合與陸面反饋機制，在模擬登陸後之強度衰減及路徑偏折亦不如物理模型精確 (Sudharsan et al., 2025)。綜上所述，顯示出 AIWP 在現階段仍難以完全取代動力模式。

汲取前述 ECMWF 與 NOAA 等先進氣象中心之實務經驗，有助於建立未來作業的分工藍圖。AI 模型之卓越運算效率，能承擔大規模系集成員生成、快速敏感性測試，以及資料同化系統中之背景場估計與產品後處理等任務，這優點應充分利用。另一方面，傳統 NWP 與人類預報員則專注於需要物理深度解析、經驗判斷並承擔決策責任之核心領域，特別是對於複雜物理過程的精細模擬、極端事件強度的最終裁奪，以及將科學預報轉化為對社會具實質意義的風險警示；此等關鍵環節始終需要仰賴具備在地知識與社會責任感之人類專家來完成 (Schultz et al., 2021)，乃是氣象署預報員未來充實專業的方向。

四、方法論之挑戰與次世代演進

AIWP 在計算效率上之革命性突破，無疑為其最顯著的優勢。此種效率提升並非漸進式改良，而是數量級之跨越，所帶來之影響不僅在於以更快速度完成相同任務，更在於實現過去無法達成之應用：傳統系集預報受限於計算資源通常僅能生成 50 至 100 個成員；而藉由 AIWP 之高效能，使生成千級甚至萬級系集成員的系集預報從理論構想轉變為實際可行 (Mahesh et al., 2025)。大規模系集預報除了用以提供完整之不確定性資訊，更能有效辨識低機率但高衝擊之極端事件軌跡，對防災決策具有重大意義。然而，秉持科學誠信原則，吾人必須審慎檢視 AI 氣象預報之侷限與風險。

深入剖析 AI 模型之預報產出，一個普遍且棘手之問題浮現：預報場過度平滑化。這意味著小尺度天氣細節被抹平，另外則是極端值被系統性低估。此並非單一模型之偶發缺陷，而是當前主流訓練方法論之固有特性 (Price et al., 2024; Waqas et al., 2025)。平滑化問題有兩個相互獨立的結構性根源。其一為損失函數設計，二為影響訓練資料之解析度上限。第一代 AI 氣象模型多採用均方誤差 (MSE) 或其變體作為損失函數，MSE 之訓練目標為最小化預測值與真實值差異之平方和，此隱含一危

險假設：所有預報誤差之權重相同，而地區性跟極端值的誤差則被消弭。換言之，AI 模型學習之最優策略為給出接近氣候平均值之結果，涵蓋極端天氣發生的頻率偏低；從統計優化角度而言，AI 模型傾向於犧牲極端值準確度以換取整體誤差之降低。

此種平滑化效應在視覺上表現為預報場之「模糊感」(blurring)，相較於高解析度 NWP 或衛星觀測，AI 繪製之鋒面不夠銳利、颱風眼牆結構鬆散、對流邊界過於圓滑。此種模糊直接導致極端值低報。在颶風強度預測上，基於 MSE 訓練之 AIWP 強度普遍低估約 20% (Price et al., 2024)。最新作業化評測顯示，AIWP 之系統性偏差導致颱風因過度平滑而難以維持其動力結構 (Shi et al., 2025)。在極端降水事件中此問題尤為突出，FourCastNet 在極端降水預報上之低報幅度可達 35%，遠高於傳統 IFS 系統之 15% (Ben-Bouallegue et al., 2023)。此一問題不僅影響預報準確度，更可能導致災害預警系統低報風險，產生嚴重之社會後果。

另一方面，AI 模型之表現上限受限於其訓練資料之品質。Pangu-Weather 研究團隊明確指出，模型之所以難以準確預報極端強度，主要歸因於訓練資料 ERA5 之水平解析度 (25 公里) 過於粗糙，導致模型

天生缺乏學習極端強度之「正確答案」，無法解析熱帶氣旋精細之內核結構 (Bi et al., 2023)。資料中存在系統性偏差或缺陷，模型不僅會學習並忠實複製此等特性，甚至可能放大之。這樣的困境會形成惡性循環：因解析度不足而無法準確表徵極端事件，導致 AI 模型學習錯誤或不完整之物理關係，進而在實際預報中系統性低估極端強度。即使投入更多計算資源訓練更大之模型，若訓練資料之根本限制未獲解決，預報能力之提升亦將觸及天花板。此問題在熱帶氣旋強度預報上尤為顯著。

與訓練資料品質相關之另一重大挑戰，乃氣候不變性假設 (climate stationarity assumption) 之侷限。AI 模型在訓練資料時空範圍內表現良好，但若應用於訓練範圍之外時，其預報能力是否能夠維持？氣候變遷正在改變大氣統計特性——極端事件頻率與強度持續增加，降水型態出現顯著變化，某些地區之氣候特徵正在經歷前所未有之轉變。當未來氣候狀態與過去歷史資料顯著不同時，基於歷史資料訓練之 AI 模型能否維持其預報水準？此已非理論推測，而是迫切之現實擔憂。舉例來說，當將 GraphCast 等模型應用於其訓練資料中未曾見過之破紀錄極端事件時，其預報能力出現可量化之顯著衰退——不僅誤差大幅飆升，對極端強度之掌握度甚至輸給

傳統數值模式。突顯了 AI 模型在面對氣候變遷所導致之分布外（out-of-distribution）情境時，存在顯著脆弱性（Zhang et al., 2025）。這對於長期氣候預測與氣候變遷調適規劃具有深遠影響，因此類應用恰需預測超出歷史經驗範圍之未來情境。

AI 模型面臨之第二大挑戰在於物理一致性（physical consistency）之缺失。當前主流 AIWP 多採純資料驅動架構，並未顯式納入物理守恆定律（如質量守恆、能量守恆、動量守恆），其物理合理性完全依賴於訓練資料（如 ERA5 再分析資料）本身的物理一致性。這種間接學習策略在例行作業情況下運作良好，但當模型遭遇訓練資料分布以外之極端情境時，即可能產生物理上不合理之預測（Yuval and O'Gorman, 2023）。Kochkov et al. (2024) 的研究顯示當給予 GraphCast 極端之初始場時，模型在前幾個預報時步即出現明顯違背熱力學第一定律的現象——在無外部強迫情況下，大氣總能量增加約 2%。這種違反能量守恆的結果，證實純 AI 模型在極端條件下之不穩定性，對即時決策構成較高潛在風險。儘管傳統預報也有類似的風險，相較之下，NWP 的物理偏差多屬系統性且漸進可預期。ECMWF AIFS 1.1.0 曾嘗試透過引入「物理約束層」（Physics-constrained

Layer）進行修正，例如利用 ReLU 等激活函數與診斷輸出層，強制預報變數符合非負性（Non-negativity）等物理限制，雖能消除了作業上困擾已久的「負降雨」現象，並大幅改善降雨結構的物理合理性（Moldovan et al., 2025），但本質上仍屬事後約束—只能在模型輸出端強制變數範圍。

與物理一致性問題緊密相連者，乃 AI 模型之黑盒子特性（black-box nature）。當傳統 NWP 預報出現異常結果，氣象學家可追溯動力方程之計算過程進行診斷，辨識問題源於初始場誤差、物理參數化方案缺陷，抑或數值不穩定。然而，當 AI 模型預測颱風將在 48 小時後突然轉向時，吾人往往難以解析其背後之推理邏輯。模型內部數億參數如何協同運作以產生此一預報？哪些大氣特徵對此決策貢獻最大？此等問題在當前深度學習架構下難以獲得明確解答。此種「不可解釋性」在學術探索中或可被容忍，但在分秒必爭之作業化預報與防災決策中，卻會構成嚴重之信任障礙（Schultz et al., 2021）。預報員需要理解模型為何做出某項預測，方能評估其可信度並決定是否採納；決策者需要知曉預報之不確定性來源，才能制定適當之應變策略。黑盒子特性使此等需求難以滿足。

最後，方法論之演進亦涉及對 AI 系集預報架構之重新審視。當前部分作業化測

試或部分氣象機構採用的初期架構，多採取將「基於 MSE 訓練之決定性模型」搭配「初始場擾動」之策略。這樣的系集架構雖能透過擾動生成多個預報成員，但受限於基礎模型在訓練階段即被優化為傾向輸出氣候平均態，導致在面臨極端天氣時，仍難以擺脫過度平滑之動力特徵，將傾向於系集離散度不足以涵蓋真實風險之現象。此種「偽系集」(pseudo-ensemble)雖在數量上提供多個預報成員，但在本質上難以真正捕捉大氣系統之內在不確定性。

因此，研究發展的重心正逐漸轉向基於連續等級概率分數 (Continuous Ranked Probability Score, CRPS) 訓練之「原生機率系集系統」(native probabilistic ensemble systems)。與傳統決定性指標不同，CRPS 能同時量化預報分布之準確度 (accuracy) 與銳利度 (sharpness)。將 CRPS 導入作為損失函數，意味模型訓練目標從「單點數值之逼近」轉型為「機率分布之建構」。ECMWF 已開始導入基於 CRPS 之損失函數進行 AIFS 訓練，最新進展顯示雖然此策略可能使預測誤差 (如均方根誤差) 略微上升，卻能顯著提升捕捉極端事件之能力 (Lang et al., 2024)。Google DeepMind 團隊開發之預報生成網路 FGN (Forecasting Generation Network; Alet et al., 2025) 透過學習到之模型擾動 (learned

model perturbations) 結合受約束之模型結構，直接針對 CRPS 進行優化，成功捕捉極具挑戰性之聯合空間結構 (joint spatial structure)，克服過去機率預報難以兼顧空間連貫性之難題。實證顯示，FGN 在熱帶氣旋路徑及多項機率指標上，明顯超越包括 GraphCast 與 Pangu-Weather 在內之現有主流模型，確立「原生機率系集」作為下一代 AI 氣象建模之主流方向，亦標誌著 AIWP 技術正經歷從「決定性模型之擾動模擬」邁向「真實機率預報」之關鍵轉型；此一轉變不僅為技術層面之進步，更代表對預報本質認知之深化——承認大氣系統之內在隨機性，並將此不確定性如實傳達予使用者。

五、未來研究方向與作業發展策略

展望 AIWP 之未來發展，吾人不應期待單一模型能解決所有問題，而應採取複合式、風險導向之發展策略。關鍵觀點在於：純資料驅動 AI 模型在物理一致性與極端事件預報上存在固有缺陷，恰為傳統 NWP 之強項。此互補特性清晰的指出，未來技術突破將來自於 AI 與物理方法之深度融合。

目前學術界跟業界都在探索之解決方案，大致可歸納為三條技術路線。第一條路線為物理資訊神經網路 (Physics-

Informed Neural Networks, PINNs), 其核心概念係將物理守恆定律引入深度學習的優化程序中, 透過在損失函數中嵌入物理約束, 確保模型預測嚴格遵循物理控制方程 (Karniadakis et al., 2021)。此方法將質量、能量及動量守恆律轉化為損失函數中的懲罰項, 使得模型在最小化預報誤差的同時, 也須滿足流體力學基本方程。儘管此舉增加訓練複雜度, 卻能顯著改善長時效預報之穩定性與物理合理性。第二條路線為模組化混合架構 (modular hybrid architecture), NeuralGCM 可做為代表; 此類系統保留傳統 NWP 的動力核心, 以確保流體運動之基本守恆律, 而將運算成本高昂且最具不確定性的次網格物理參數化過程 (如輻射傳輸、雲微物理及邊界層湍流), 交由神經網路構建之代理模型 (Surrogate Models) 進行端到端學習。此種設計不僅發揮了 AI 捕捉複雜非線性物理反饋的優勢, 更透過動力核心的約束, 同時避免純 AI 模型在長時效積分中可能出現之漂移與能量不守恆問題。第三條路線為 ECMWF 正在實驗的光譜微調策略 (spectral nudging): 它的核心機制係在傳統 IFS 模式運行過程中, 將 AI 預測之大尺度環流場作為背景約束, 透過光譜濾波技術引導 IFS 的大尺度結構, 使其與 AI 預報趨勢保持一致。在此架構下, 其細微的次網

格過程 (如地形降水、重力波傳播等) 則保留傳統物理參數化方案進行精密計算。初步測試顯示此種混合預報在保持 IFS 物理一致性之同時, 顯著改善大尺度環流預報技能 (Polichtchouk et al., 2025)。

第四節述及, 解決由於 MSE 損失函數之統計特性造成的平滑化問題, 須從「提供單點估計」之決定性預報, 轉向「提供機率分布」之生成式機率預報 (generative probabilistic forecasting)。而擴散模型 (Diffusion Models) 為此領域近年最成功之架構。前文提及 CorrDiff 在颱風降尺度上之卓越表現 (Mardani et al., 2025), 若將此技術擴展至全球預報層次, 系統即可生成數千個「物理合理、統計一致、但細節各異」之預報場——每一個均代表未來之一種真實可能性。這是面對極端天氣應有之態度: 並非告訴決策者「將會如何」, 而是告訴其「可能如何」, 並量化每種可能性之發生機率。機率資訊對風險管理極其重要, 使決策者能根據風險容忍度制定適當應變策略, 而非被迫依據單一預報場景做出二元決策。

對臺灣而言, AI 氣象技術之發展既無法亦無需與 ECMWF 或 NOAA 等國際氣象權威機構在「全球基礎模型」之算力競賽中正面交鋒——此類競賽需要龐大之計算資源、海量之訓練資料, 以及長期之技

術累積。因著 AI 技術大幅降低發展「區域化、專門化」預報能力之門檻，成為臺灣氣象發展的一個轉機，策略核心應是為建立一套以國際先進模型為基底、再輔以在地校正技術之複合體系。臺灣可充分利用國際先進開源 AI 模型作為預報基礎，集中資源發展針對臺灣複雜地形與特殊氣候現象（如梅雨鋒面與颱風之交互作用、地形降水之精細預報、局地對流系統之快速發展）之後處理與降尺度技術，避免重複投資全球模型開發，將有限資源聚焦於最能發揮在地優勢之領域。

為實現此一策略，建立高品質之臺灣區域再分析資料庫（Taiwan regional reanalysis database）不僅為技術需求，更應被提升為國家級數位基礎建設之高度來推動。高品質再分析資料為 AI 模型訓練之基石。臺灣擁有密集之觀測網路（包括地面氣象站、自動雨量站、雷達網、探空系統），透過先進資料同化技術，將觀測整合為高時空解析度（如水平 2 公里、每小時更新）之再分析資料集，可涵蓋過去二十年至三十年之完整氣候變化週期，做為 AI 模型提供學習臺灣特殊天氣系統之訓練樣本。此資料庫不僅可用於 AI 模型訓練，亦可支持氣候變遷研究、災害風險評估、水資源管理等多元應用，具有長遠之決策價值。同時應著力發展高解析度降尺度技術，使

全球模型之粗解析度預報能有效轉化為捕捉臺灣地形細節之高解析度預報。生成式 AI（如擴散模型）在此領域已展現顯著潛力，未來研究應探索如何結合物理約束與生成式模型，產生既符合大尺度動力一致性、又能精確描繪局地地形效應之降尺度產品。

預報之終極價值不僅在於準確預測溫度、氣壓、風速等物理量，更在於預測此等天氣條件對社會運作之實質影響。隨著 AIWP 的作業化與高效能預報能力日益成熟，其高速運算與豐富的不確定性資訊，為後續的決策轉譯應用提供了堅實基礎。AI 在未來更可進一步扮演決策轉譯者（decision translator）的關鍵角色：透過大型語言模型（Large Language Models, LLMs）的轉譯能力整合多源資料——除了氣象觀測，再納入社會經濟數據、基礎設施分布、人口流動及歷史災情記錄，從而提供真正「有感」的預報（Impact-Based Forecasting；Waqas et al., 2025）。氣象署長期與衛福部、勞動部、農業部、水利署及水保局等單位合作，提供高溫健康風險、農業影響及防洪支援等跨域服務，惟此類工作需投入大量人力進行資料整合與轉譯，而 LLMs 的導入可有效降低跨域門檻、提升轉譯效率，進一步將既有專業模式輸出快速轉化為個人化行動建議。此種以使用者為中心的有

感預報，正協助氣象服務從資訊提供逐步進化為主動決策支援。

在資源分配方面，臺灣應採取務實之優先順序策略。首要投資應為高品質觀測網路之維護與提升，因無論 AIWP 或 NWP 皆高度依賴準確之初始場資料。再者為投資人才培育，培養同時具備氣象物理知識與 AI 技術能力之跨領域人才，能架起氣象學家與電腦科學家之溝通橋樑，確保技術發展符合實際需求。第三應為國際合作，透過參與國際 AI 氣象研究計畫、與先進氣象中心建立技術交流管道，即時掌握最新技術動態並取得訓練資源。最後，建立開放之資料與模型共享平台，促進學術界、作業界與產業界之協作，加速技術從研究原型到作業應用之轉化。

六、結論：共生而非替代

回顧 2023 年至今的發展歷程，人工智慧氣象預報技術演進軌跡在 3 年間迅速完成從「技術可行性驗證」到「作業化初步整合」的跨躍。然而，透過本研究之系統性回顧與個案分析，顯見純資料驅動模型與物理數值模式之間的關係，並非「AI 取代傳統預報」之零和對立，而是「深度整合」的推進。這場變革不僅是運算效率的革新，更是一場關於預報本質、作業流程與風險治理的全面轉型。

值得高度關注的是，AI 預報技術之普及已顯著降低了過往由高昂算力資源與硬體設施所構築的技術壁壘。週邊國家正藉由引進國際開源基礎模型與雲端計算資源，直接跨越傳統數值天氣預報在資料同化與模式研發上的長期技術門檻，實現技術的跳躍演進。這種趨勢有利於區域競爭者能在短期內掌握極具競爭力的預報能力，甚至在特定極端事件預測上挑戰臺灣之固有優勢。面對此種區域性的技術重組與競爭壓力，針對臺灣氣象作業之策略定位，本研究強調臺灣不應僅侷限於全球基礎模型的追隨者，而應利用自身作為「複雜地形實驗場」的獨特地理特性，建立具主動權之複合式預報體系。在國際上的合作策略也應重新審視：從單向技術輸出轉向雙向技術交流，並在 AI 應用之某些面向保持開放學習態度。

在上述策略架構下，本研究透過建立系統性分類框架，論述 AI 氣象模型之多樣性與互補性，從即時預報至次季節預測，從端到端系統至混合架構，乃至從全球模式至區域降尺度，不同 AI 技術正精準對接不同預報需求。對技術限制之誠實剖析乃本研究另一核心貢獻：預報場之平滑化、極端事件之低估、物理一致性之隱憂以及黑盒子之不可解釋性，皆根植於當前 AI 方法論之固有特性。指認此等限制並非否定

AI 之價值，而是為建立正確預期管理，並確立其在預報作業中之合理定位。展望未來，AI 與物理混合架構為科學界的共識，生成式機率預報為技術演進的趨勢，而臺灣突圍的方向則在於區域化與在地知識。

具體而言，高品質「臺灣區域再分析資料庫」的建置，不應僅被視為技術研發的需求，更應提升至國家級數位基礎建設的高度。這項建設將賦予臺灣在西北太平洋熱帶氣旋的專門研究與局地對流的精細預報中，將「全球 AIWP 空間廣度」與「在地物理動力深度」進行深度整合之戰略能力，進而實現更具韌性的氣候風險治理。

AIWP 的普及並非取代人類預報員，而是驅動其職能重心由資料處理轉向風險溝通。預報員之專業價值將聚焦於極端情境下的物理診斷與決策支援。為因應此趨勢，人才培育體系需啟動結構性改革：大氣科學教育應將統計學習與深度學習納入核心必修，物理教學重點則應從傳統的解析解推導，轉向對動力機制「物理直覺」之培養。透過訓練具備模型評估能力與大氣動力背景的「雙語人才」，方能確保臺灣在技術浪潮中，建構一個由 AI 賦能、物理約束且以風險決策為核心的共生預報體系，此為本回顧研究對氣象科學未來發展之核心願景。

致謝

作者感謝中央氣象署提供觀測資料與作業經驗分享，以及 ECMWF、Google DeepMind、NVIDIA 等機構對 AI 氣象預報技術開源之貢獻。本研究獲益於匿名審查委員之寶貴意見，特此致謝。

參考文獻

- Alet, F., Price, I., El-Kadi, A., Masters, D., Markou, S., Andersson, T. R., Stott, J., Lam, R., Willson, M., Sánchez-González, Á., & Battaglia, P. (2025). Skillful joint probabilistic weather forecasting from marginals. *arXiv preprint*, <https://doi.org/10.48550/arXiv.2506.10772>
- Arcomano, T., Szunyogh, I., Pathak, J., Wikner, A., Hunt, B. R., & Ott, E. (2020). A machine learning-based global atmospheric forecast model. *Geophysical Research Letters*, 47(9), e2020GL087776. <https://doi.org/10.1029/2020GL087776>
- Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55. <https://doi.org/10.1038/nature14956>

- Ben-Bouallegue, Z., Clare, M. C. A., Magnusson, L., Gascon, E., Maier-Gerber, M., Janousek, M., ... & Pappenberger, F. (2023). The rise of data-driven weather forecasting: A first statistical assessment of the machine learning models Pangu-Weather and GraphCast. *Geophysical Research Letters*, 51(2), e2023GL105966. <https://doi.org/10.1029/2023GL105966>
- Ben Bouallègue, Z., and Coauthors, 2024: The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning-Based Weather Forecasts in an Operational-Like Context. *Bull. Amer. Meteor. Soc.*, 105, E864–E883, <https://doi.org/10.1175/BAMS-D-23-0162.1>
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619, 533–538. <https://doi.org/10.1038/s41586-023-06185-3>
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., ... Perdikaris, P. (2025). A foundation model for the Earth system. *Nature*, 641, 1180–1187. <https://doi.org/10.1038/s41586-025-09005-y>
- Brenowitz, N. D., Cohen, Y., Pathak, J., Mahesh, A., Bonev, B., Kurth, T., Durran, D. R., Harrington, P., & Pritchard, M. S. (2025). A practical probabilistic benchmark for AI weather models. *Geophysical Research Letters*, 52, e2024GL113656. <https://doi.org/10.1029/2024GL113656>
- ECMWF, 2025: AIFS Single and Ensemble: Operational performance and evaluation. *ECMWF Technical Memorandum*, 185, <https://doi.org/10.21957/hglz-pe65>
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3, 422–440. <https://doi.org/10.1038/s42254-021-00314-5>
- Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., & Hoyer, S. (2024). Neural general circulation

- models for weather and climate. *Nature*, 632(8027), 1060–1066. <https://doi.org/10.1038/s41586-024-07744-y>
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Al-Shedivat, S. A. K., Rezende, D. J., & Battaglia, D. C. (2023). Learning skillful medium-range global weather forecasting with Graph Neural Networks. *Science*, 382(6677), 1416–1421. <https://doi.org/10.1126/science.adi2336>
- Lang, S., Bauer, P., Geer, A., Arcucci, R., Bauer, M., Ben-Bouallègue, Z., ... & Wilhelm, T. (2024). AIFS—ECMWF's data-driven forecasting system. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2406.01465>
- Li, H., Chen, L., Zhong, X., Wu, J., Chen, D., Xie, S.-P., Chao, Q., Lin, C., Hu, Z., Lu, B., & Qi, Y. (2024). A machine learning model that outperforms conventional global subseasonal forecast models. *Nature Communications*, 15, 6936. <https://doi.org/10.1038/s41467-024-50714-1>
- Mahesh, A., Collins, W. D., Bonev, B., Brenowitz, N. D., Cohen, Y., Elms, J., Harrington, P., Kashinath, K., Kurth, T., North, J., O'Brien, T., Pritchard, M., Pruitt, D., Risser, M., Subramanian, S., & Willard, J. (2025). Huge ensembles – Part 1: Design of ensemble weather forecasts using spherical Fourier neural operators. *Geoscientific Model Development*, 18(17), 5575–5603. <https://doi.org/10.5194/gmd-18-5575-2025>
- Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., & Pritchard, M. (2025). Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Communications Earth & Environment*, 6, 124. <https://doi.org/10.1038/s43247-025-02042-5>
- Moldovan, G., Pinnington, E., Prieto Nemesio, A., Lang, S., Ben Bouallègue, Z., Dramsch, J., Alexe, M., Santa Cruz, M., Hahner, S., Cook, H., Theissen, H., Clare, M., O'Brien, C., Polster, J., Magnusson, L., Mertes, G., Pinault, F., Raoult, B., de Rosnay, P., Forbes, R., & Chantry, M. (2025). AIFS 1.1.0: An update to ECMWF's machine-learned weather forecast model AIFS.

- EGUsphere*. <https://doi.org/10.5194/egusphere-2025-4716>
- National Weather Service. (2024). *Public Information Statement: Soliciting Comments on the Experimental Artificial Intelligence Global Forecast System (AIGFS)*. NOAA. https://www.weather.gov/notification/pns24-aigfs_experimental
- National Weather Service. (2025). *Service Change Notice 25-89: Operational Implementation of the Artificial Intelligence Global Forecast System (AIGFS) Version 1.0*. NOAA. https://www.weather.gov/notification/scn25-89_aigfs_v1
- Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., Hassanzadeh, P., Kashinath, K., & Anandkumar, A. (2022). FourCastNet: Accelerating global high-resolution weather forecasting with adaptive Fourier neural operators. *arXiv preprint*, <https://doi.org/10.48550/arXiv.2202.11214>
- Polichtchouk, I., Clare, M., Chantry, M., Maier-Gerber, M., & Lang, S. (2025). Hybrid forecasting: Nudging large scales of the IFS to the deterministic AIFS. *ECMWF Newsletter*, 184, 14–19.
- Price, I., Hoyer, S., Bentley, R., & Coauthors. (2024). GenCast: Diffusion-based ensemble forecasting for medium-range weather. *arXiv preprint*, <https://doi.org/10.48550/arXiv.2312.15796>
- Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., ... Willson, M. (2025). Probabilistic weather forecasting with machine learning. *Nature*, 637(8044), 84–90. <https://doi.org/10.1038/s41586-024-08252-9>
- Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878), 672–677. <https://doi.org/10.1038/s41586-021-03854-z>
- Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., & Stadtler, S.

- (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200097. <https://doi.org/10.1098/rsta.2020.0097>
- Shi, Y., Hu, R., Wu, N., Zhang, H., Liu, X., Zeng, Z., Zhu, J., Han, P., Luo, C., Zhang, H., He, J., & Shi, X. (2025). Comparison of AI and NWP models in operational severe weather forecasting: A study on tropical cyclone predictions. *JGR: Machine Learning and Computation*. <https://doi.org/10.1029/2024JH000481>
- Sønderby, C. K., Espenholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., ... & Kalchbrenner, N. (2020). MetNet: A Neural Weather Model for Precipitation Forecasting. *arXiv preprint arXiv:2003.12140*. <https://doi.org/10.48550/arXiv.2003.12140>
- Sudharsan, N., Singh, M., Talukdar, S., Mohanty, S., Kamath, H., Osuri, K. K., Dashtian, H., Young, M., Yang, Z.-L., Dawson, C., Leung, L. R., Gopalakrishnan, S., Mehra, A., Tallapragada, V., & Niyogi, D. (2025). Enhancing near real time AI-NWP hurricane forecasts: Improving explainability and performance through physics-based models and land surface feedback. *arXiv*. <https://doi.org/10.48550/arXiv.2502.01797>
- Tabas, S. S., Wang, J., Li, W., Row, M., Zhang, Z., Zhu, L., Peng, J., & Carley, J. R. (2024). GFS-powered machine learning weather prediction: A comparative study on training GraphCast with NOAA's GDAS data for global weather forecasts. *arXiv*. <https://doi.org/10.48550/arXiv.2407.03960>
- Vitart, F., & Robertson, A. W. (Eds.). (2018). *Sub-seasonal to seasonal prediction: The gap between weather and climate*. Elsevier. <https://doi.org/10.1016/C2016-0-01993-9>
- Waqas, A., Heaney, C., Yan, K., Agarwal, S., Varela, J., & Kaouri, K. (2025). Data-driven weather prediction: Opportunities and challenges. *Artificial Intelligence Review*, 58, 31.

- <https://doi.org/10.1007/s10462-024-11110-3>
- Yuval, J., & O'Gorman, P. A. (2023). Neural-network parameterization of subgrid-scale forcing does not require high-resolution simulation data. *Geophysical Research Letters*, 50(16), e2023GL104768. <https://doi.org/10.1029/2023MS003606>
- Zhang, Z., Fischer, E. M., Zscheischler, J., & Engelke, S. (2025). Numerical models outperform AI weather forecasts of record-breaking extremes. *arXiv preprint arXiv:2508.15724*. <https://doi.org/10.48550/arXiv.2508.15724>

AI-Driven Weather Prediction: From Technical Innovation to Operational Decision Symbiosis

Hsiao-Wei Lai^{1*}, Shih-Yu Wang², Jing-Shan Hong³, Chi Pan³,

Meng-Shih Chen³, Han-Chun Wang¹, Chih-Hsien Tsai¹

¹ Environmental Simulation Co., Ltd.

² Research Center for Environmental Changes, Academia Sinica,

National Taiwan University, Taipei, Taiwan

³ Central Weather Administration, Taipei, Taiwan

(manuscript received 21 October 2025; in final form 6 May 2026)

Abstract

In 2023, Google DeepMind's GraphCast model, which utilizes a Graph Neural Network (GNN) architecture for Artificial Intelligence Weather Prediction (AIWP), demonstrated superior forecasting performance compared to traditional ECMWF numerical models. It surpassed the European Centre for Medium-Range Weather Forecasts (ECMWF) standards across 90% of verification variables, while reducing computational time by three orders of magnitude. This study reviews the paradigm shift in meteorological forecasting driven by AI over the past three years. We propose a classification framework for AIWP based on temporal scales and operational roles, covering nowcasting (0–6 h), medium-range forecasting (1–10 d), and sub-seasonal to seasonal (S2S) prediction. The review traces the technical evolution from end-to-end models to hybrid architectures, and examines how next-generation systems such as Aurora and GenCast use foundation models and generative AI to overcome traditional constraints. Within the S2S domain, the FuXi-S2S model extends the effective lead time of the Madden-Julian Oscillation (MJO) to 36 days. By employing a flow-dependent perturbation scheme, it demonstrates superior early warning capabilities for extreme weather events. However, current AIWP systems face inherent limitations due to Mean Squared Error (MSE) loss functions and a lack of physical conservation constraints. These factors result in spectral smoothing, physical inconsistency, and a systematic underestimation of extreme events. Such deficiencies are evidenced by a 20–50%

underestimation in tropical cyclone intensity and a 35% bias in extreme precipitation. Consequently, the shift toward physics–AI hybrid architectures has become inevitable. For Taiwan's meteorological operations, this study recommends a strategy of "international foundation models with localized calibration". Priorities include building a high-quality Taiwan regional reanalysis database, developing high-resolution downscaling techniques, and extending these capabilities to climate risk governance.

Keywords: AI Weather Prediction, Deep Learning, Climate Prediction, Extreme Weather, Downscaling Techniques. doi: 10.53106/025400022026065401002

*Corresponding author: Hsiao-Wei Lai (E-mail: hiaubi@simenvi.com.tw)