

Central Weather Administration

cwa.gov.tw



大型機器學習天氣預報模型之應用經驗

112年天氣分析與預報研討會

2023 Conference on Weather Analysis and Forecasting

A7 人工智慧之氣象應用 (I)

劉正欽、徐驊、彭順台、陳得松、張保亮、蕭
玲鳳、陳新淦、陳品璇、沈彥志、蔡金成


2024/09/04



 前言-基本資訊簡介

 (大型)全球AI/ML天氣模型之應用經驗

 (大型)全球AI/ML天氣模型之評估

 結論與建議



小八貓說：也就是說，是...的意思嗎！？

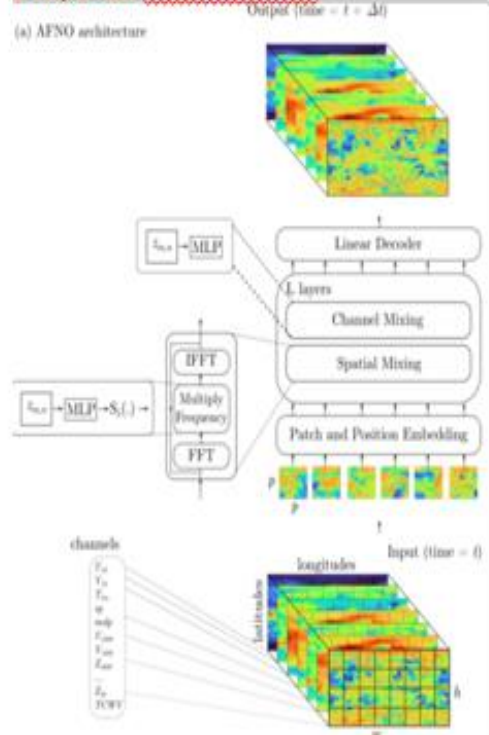
聽完這個分享後，
也就是說，是可以知道怎麼選擇損失函數的意思嗎！？
也就是說，是可以知道怎麼選擇優化器的意思嗎！？
也就是說，是可以知道怎麼建立大型模型的意思嗎！？



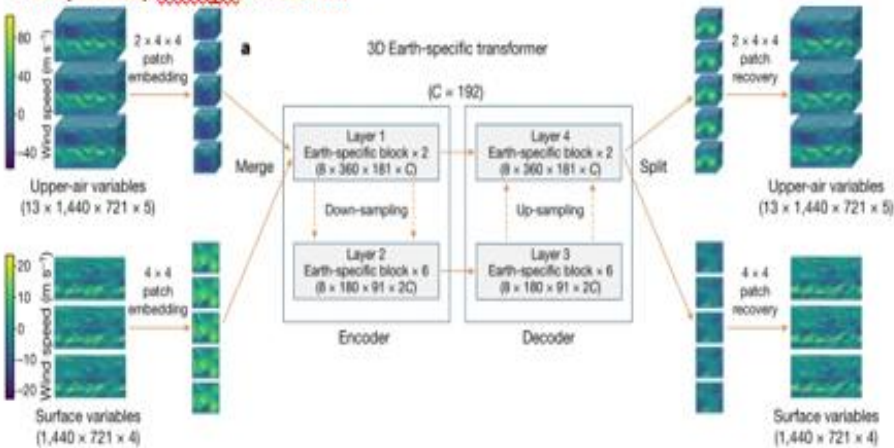
那可以知道到什麼？？
繼續聽下去~

- 前言-基本資訊簡介
- 去(2023)年的天氣分析研討會

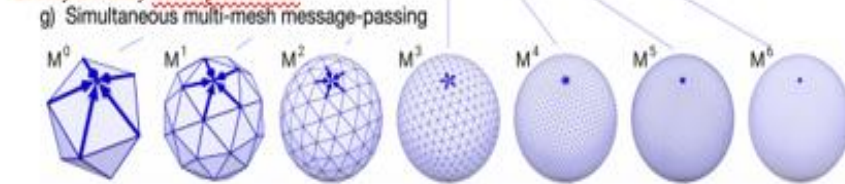
Feb, 2022 FourCastNet



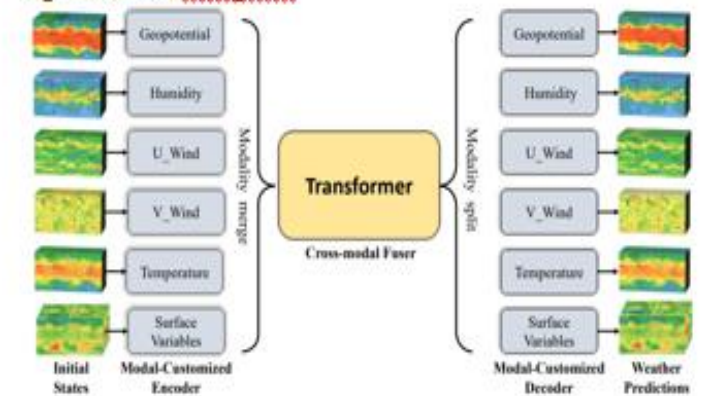
Nov, 2022, Pangu-weather



Dec, 2022, GraphCast

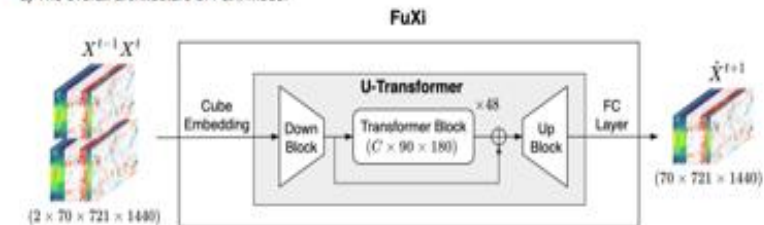


April, 2023, FengWu



June, 2023, Fuxi

a) The overall architecture of FuXi model



	FourCastNet	Pangu-Weather	GraphCast
Forecast scores	Comparable to IFS	Better than IFS	Better than IFS

Thomas Rieutord, (Mar, 2023), A review of the recent papers on fully data-driven NWP with AI

- 前言-基本資訊簡介
- 今(2024)年的天氣分析研討會

監測
0小時

即時預報
0-1小時

極短期
0-6小時

短期
1-3日

短中期
3-7日

中期
7-15日

中長期(月)
15-30日

季
30-90日

年、代、際
~

CAPN
(Ashesh et al., 2022)
MetNet-3
(Andrychowicz et al., 2023)
Nowcasnet
(Zhang et al., 2023)
Prediff
(Gao et al., 2023)
Diffcast
(Yu et al., 2023)
SCAN
(Pan et al., 2024)
CastCast
(Gong et al., 2024)

AIFS



- AI在大氣科學的應用已快速發展，包含非大氣科學領域也一同合作
 - 產官學研協力及跨域合作的重要性明顯上升 (Bauer et al., 2023; Bauer 2024; Eyring et al., 2024)
- 一同重視AI應用於氣象科學議題
 - 均衡發展，適宜的資歷及演算法解決問題。
- 資料的重要性
 - 何處有接近AI Ready Data 的資料
- 軟硬體支援
 - 執行與測試的理想環境(天地)

(Xu et al., 2024)
Zhiji
(深圳氣象局及華為, 2024)

(Microsoft, 2024)

FuXi-S2S
(Chen et al., 2023)
ChaosBench
(Nathaniel et al., 2024)

naX
(Nguyen et al., 2023)
tt-Meyer et al., 2023)
GCM
(Chkov et al., 2023)

降尺度

CorrDiff
(Mardani et al., 2023)
H4X5-MSRN
(Hsu et al., 2024)
StormCast
(Pathak et al., 2024)

- 前言-基本資訊簡介
- 氣象署的AI經驗從哪來？

1、預警產品

颱風預報

- 陣風預報
- 強度及結構分析預報
- 季預報指引
- ETQPF後處理

(強)降雨

- 短期降雨
- 整合雙偏極化雷達與雨滴譜儀觀測資料改善雷達QPE
- 衛星降雨反演
- 閃電預警應用
- 台灣地區極端因子預報

暴潮海象

- 海岸裂流監測與預警
- 沿岸異常波浪預警

地震預警

- 現地預警分析有效減少盲區
- 地震預警



2、數值模式產品優化

- 地面產品優化
- 發展區域數值模式數位孿生產品
- 氣候預報
- 偏差修正後處理
- 三至四週環流預報

3、監測應用

- 雲量辨識
- 飛機積冰
- 氣候分類
- 資料數位化資料建置
- 雷雨胞追蹤
- 目標海域全天候逐時海表溫產品

4、綠能

- 日射量預報
- 雲量預報
- 100米風及短波輻射校正
- 日太陽能預報
- 全天空照相儀進行短期日射量預報

5、健康

- 空氣品質預報
- 熱指數預報
- 氣膠光學厚度
- PM2.5濃度地面真實場產製

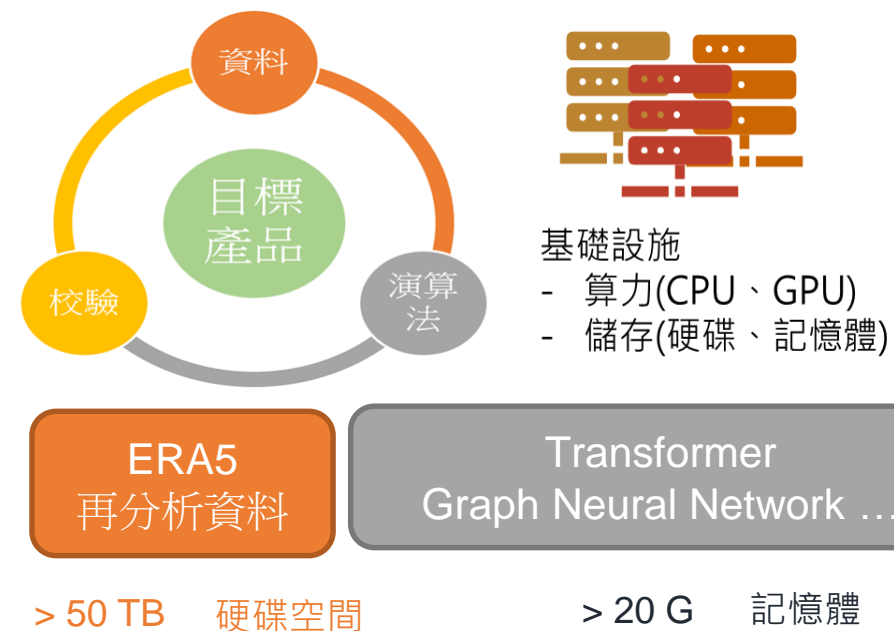
6、其他

預報海水葉綠素含量技術

- 前言-基本資訊簡介
- 大型機器學習天氣預報模型

模型名稱	訓練資料長度(年)	天氣變數		採用模型(參數量)
		垂直層	單一層	
Pangu-Weather	1979-2017 (39)	(z、t、q、u、v) ⁺	t2, 10u, 10v, mslp	Transformer (7.4億)
GraphCast	1979-2015 (37)	(z、t、q、u、v、w) [#]	t2,10u,10v,mslp,tp	圖像神經網路 (0.6億)
FengWu	1979-2015 (37)	(z、t、q、u、v) [#]	t2,u10,v10,mslp	Transformer (7.5億)
FourcastNet v2	1979-2015 (37)	(z、t、rh、u、v) ⁺	t2, 10u, 10v, sp, mslp, tcwv, 100u,100v	Transformer (4.3億)
FuXi	1979-2015 (37)	(z、t、rh、u、v) ⁺	t2,u10,v10,mslp,tp	Transformer (15.5億)
ExtremeCast	1979-2017 (39)	(z、t、q、u、v) ⁺	t2, 10u, 10v, mslp	Transformer Diffusion (7億)
Aurora	1979-2025 (37)	(z、t、q、u、v) ⁺	t2, 10u, 10v, mslp	Transformer LoRA (13億)

+50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa
 #1, 2, 3, 5, 7, 10, 20, 30, 50, 70, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000 hPa



不是一台個人電腦可以訓練!!!

研究顯示：AI 專案失敗率超過 80% · 比 AI 新創失敗率還高兩倍
作者: Evan | 發布日期: 2024 年 09 月 03 日 8:00 | 分類: AI 人工智慧, 新聞



AI 專案失敗的最大原因是核心利益關係人之間的目標不一致

- (大型)全球AI/ML天氣模型之應用經驗
- 採取行動

現況

1. 資源盤點
 - 人才
 - 資料
 - 基礎設施
2. 技術評析
 - 應用
 - 建置
3. 學習
 - 基本觀念
 - 進階觀念
 - 實作
4.



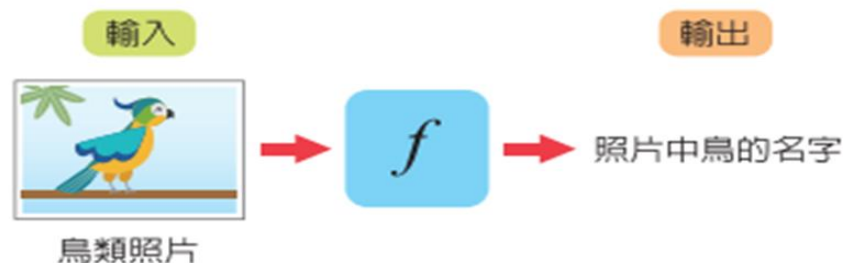
基礎設施

- 算力(CPU、GPU)
- 儲存(硬碟、記憶體)

長遠

1. 人才培育
2. 產業鏈結
3. 資料治理
4. 基礎設施建置擴充
5. 預報的需求
6. ...

這邊指的應用，是應用已建置好的模型，**只要**處理輸入的資料即可



- (大型)全球AI/ML天氣模型之應用經驗
- 如何應用

AI模型的特性-訓練需要大量資源，但應用訓練完成的模型可在一般電腦執行。
AI/ML天氣模型亦相同(Ebert-Uphoff and Hibern 2023; Wong 2023)

簡單說，一個無法用個人電腦訓練出來的模型，現在可以在一般電腦運作，不用嗎？
如果有開源，當然使用

那，想用的話，要有什麼條件？

基本需求	備案
熟悉python	認識熟悉python的人
熟悉資料格式(grib, grib2, netcdf, hdf5, hdf4, npy, pickle, zarr)	熟悉前述資料格式的人
4核心電腦，記憶體至少8G 最好是有GPU	認識有好電腦的人
看懂開源模型所需的模型輸入(讀paper)， 及懂一些深度學習套件的使用方式(不用建模~)	找到寫好的模組

- (大型)全球AI/ML天氣模型之應用經驗
- ECMWF提供的ai-models套件

<https://github.com/ecmwf-lab/ai-models>

Installation

To install the `ai-models` command, run the following command:

```
pip install ai-models
```

Available Models

Currently, four models can be installed:

```
pip install ai-models-panguweather
pip install ai-models-fourcastnet
pip install ai-models-graphcast # Install details at https://github.com/ecmwf-lab/ai-models-graphcast
pip install ai-models-fourcastnetv2
```

See [ai-models-panguweather](#), [ai-models-fourcastnet](#), [ai-models-fourcastnetv2](#) and [ai-models-graphcast](#) for more details about these models.

Running the models

To run model, make sure it has been installed, then simply run:

```
ai-models <model-name>
```

Replace `<model-name>` with the name of the specific AI model you want to run.

By default, the model will be run for a 10-day lead time (240 hours), using yesterday's 12Z analysis from ECMWF's MARS archive.

To produce a 15 days forecast, use the `--lead-time HOURS` option:

```
ai-models --lead-time 360 <model-name>
```

From a GRIB file

If you have input data in the GRIB format, you can provide the file using the `--file` option:

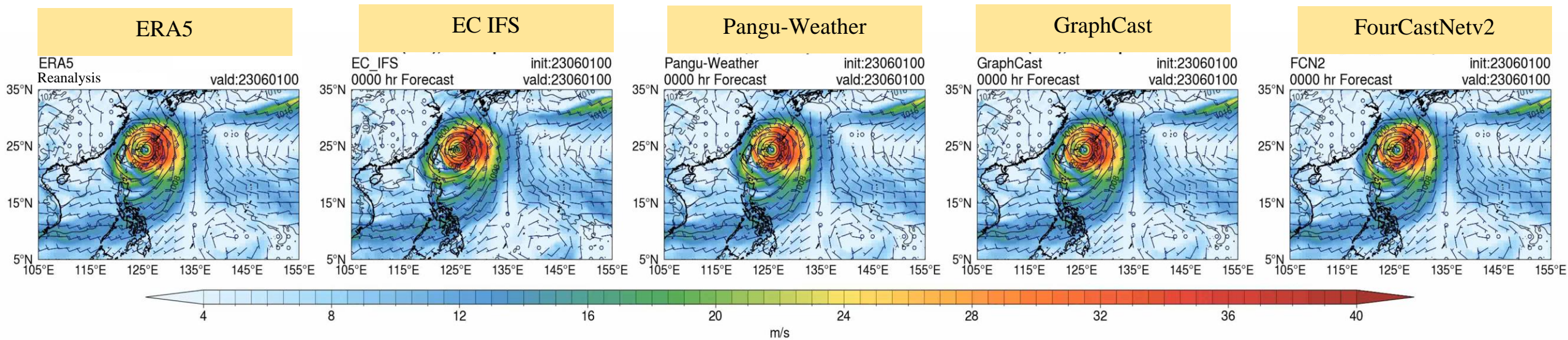
```
ai-models --file <some-grib-file> <model-name>
```

The GRIB file can contain more fields than the ones required by the model. The `ai-models` command will automatically select the necessary fields from the file.

- (大型)全球AI/ML天氣模型之應用經驗
- ECMWF提供的ai-models套件執行狀況

確認是否正確

初始時間- 2023060100 ; 初始場來源-ERA5 10天預報，瑪娃颱風離開後，三種AI天氣模型皆有反映谷超颱風的生成



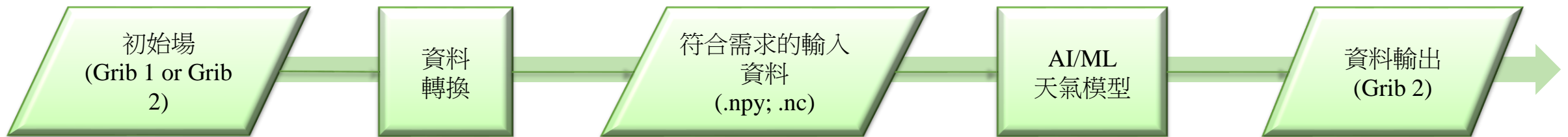
1. 慶幸沒有執行錯誤。
2. 只能跑三種(其實是四種，但當時未測試FourCastNetv1)，那FuXi跟FengWu呢？
3. 如果遇到不滿足輸入資料問題時，無法執行(如沒有重力位，只有重力位高度)。
4. 如果想應用ERA5以外的數值模式初始場(例：NCEP)，如何做？

看來要做一些客製化了....，於是就變成了業餘軟體測試工程師

- (大型)全球AI/ML天氣模型之應用經驗
- 客製化的過程

- 只能跑某幾種開源AI/ML天氣模型，那想跑其它開源AI/ML天氣模型呢？
- 如果遇到不滿足輸入資料問題時，無法執行。
- 如果想應用ERA5以外的數值模式初始場(例：NCEP)，如何做？

為了要克服左述問題，自行寫了
輸入轉換程式及推論程式



資料面可能遇到的問題

1. 不懂Grib資料格式，需先理解(讀跟寫)
2. 不懂NetCDF資料格式，需先理解(讀跟寫)
3. 天氣變數間的轉換(為了轉換成符合需求的輸入資料)
 - 如，重力位及重力位高度的轉換，相對溼度的計算等
4. 額外的資料處理，譬如Top-of-atmosphere (TOA) incident solar radiation

軟硬體面的問題

1. 記憶體不足，無法順利執行
2. 無法設定適宜的執行環境(不同版本套件的相依套件衝突，cuda版本無法支持AI/ML模型)

AI/ML天氣模型面可能會遇到的問題

1. 需先理解現行AI/ML常用套件(pytorch, tensorflow, jax, keras, onnx)，如何推論即可。
2. 同時執行多樣的AI/ML模型時，需理解GPU相關設定，若設定錯誤即無法使用GPU資源

可能還會有其它問題~~

但，如果想要執行這一些大型AI/ML天氣模型，一開始的門檻不在各位的大氣專業，絕大部分在

1. 會不會python，這個「會」包含安裝跟撰寫。
2. 有沒有基本的AI/ML知識

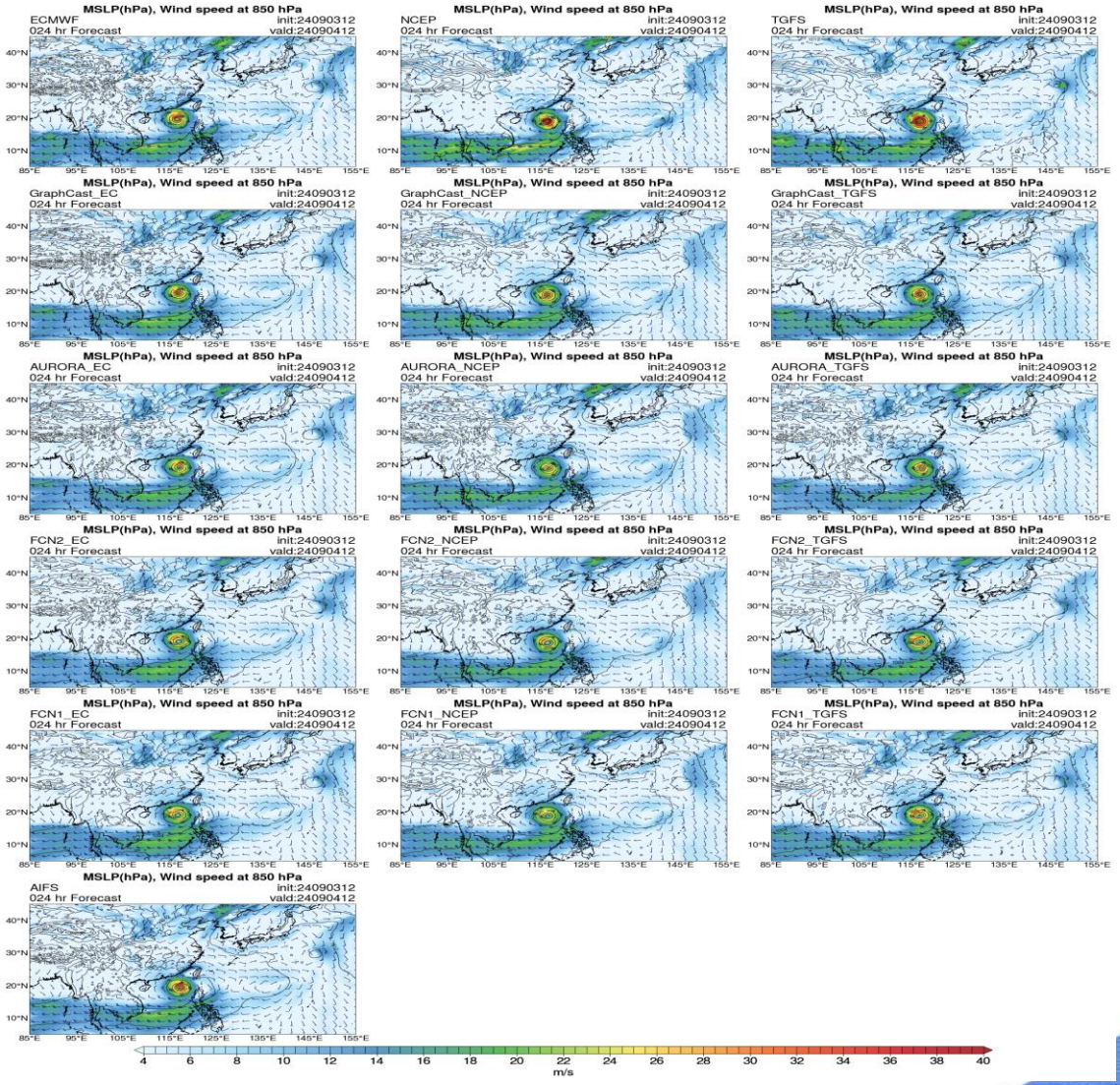
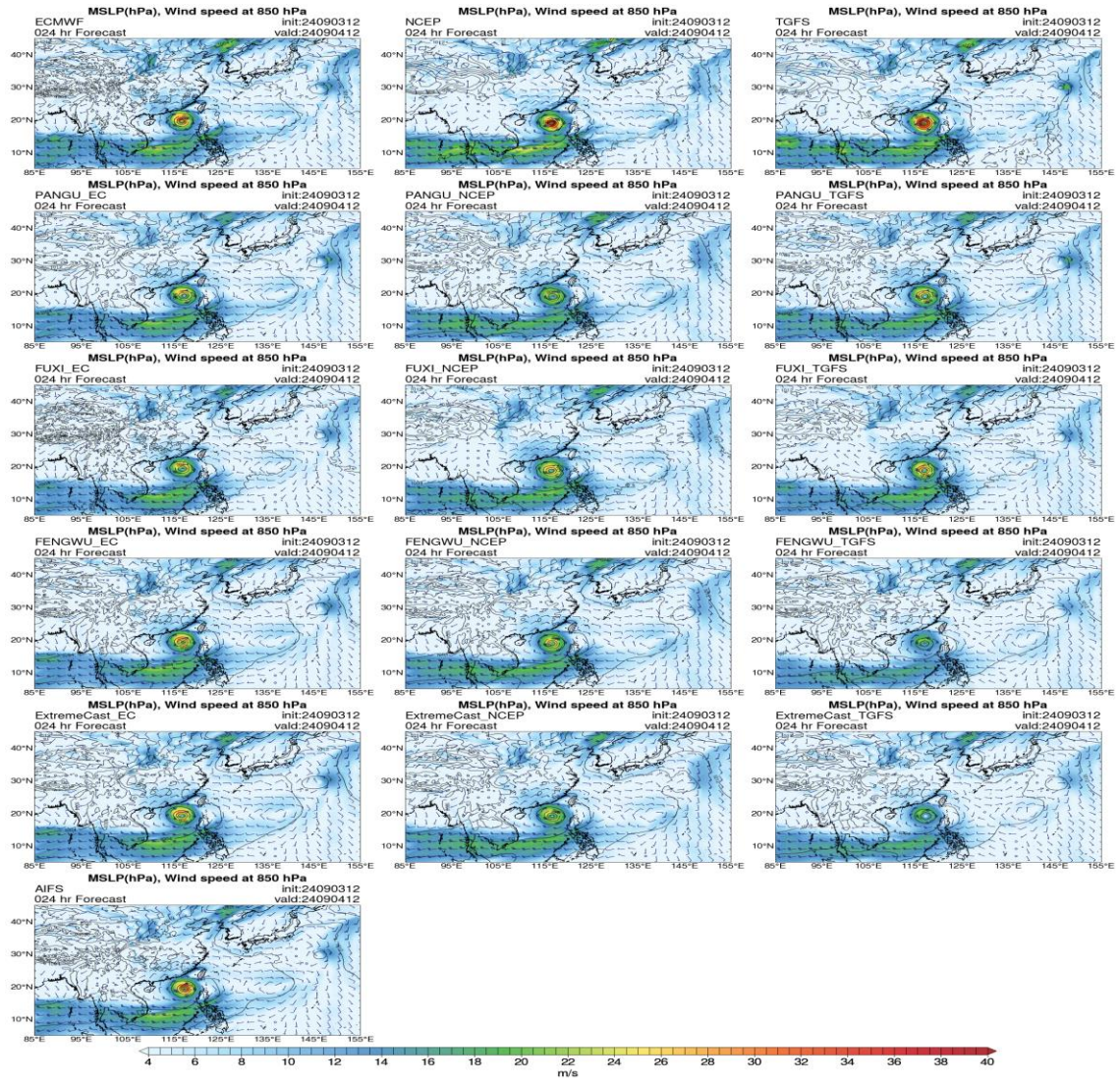
- (大型)全球AI/ML天氣模型之應用經驗
- 目前狀態

初始時間-2024090312 · 1天後(2024090412)之預報

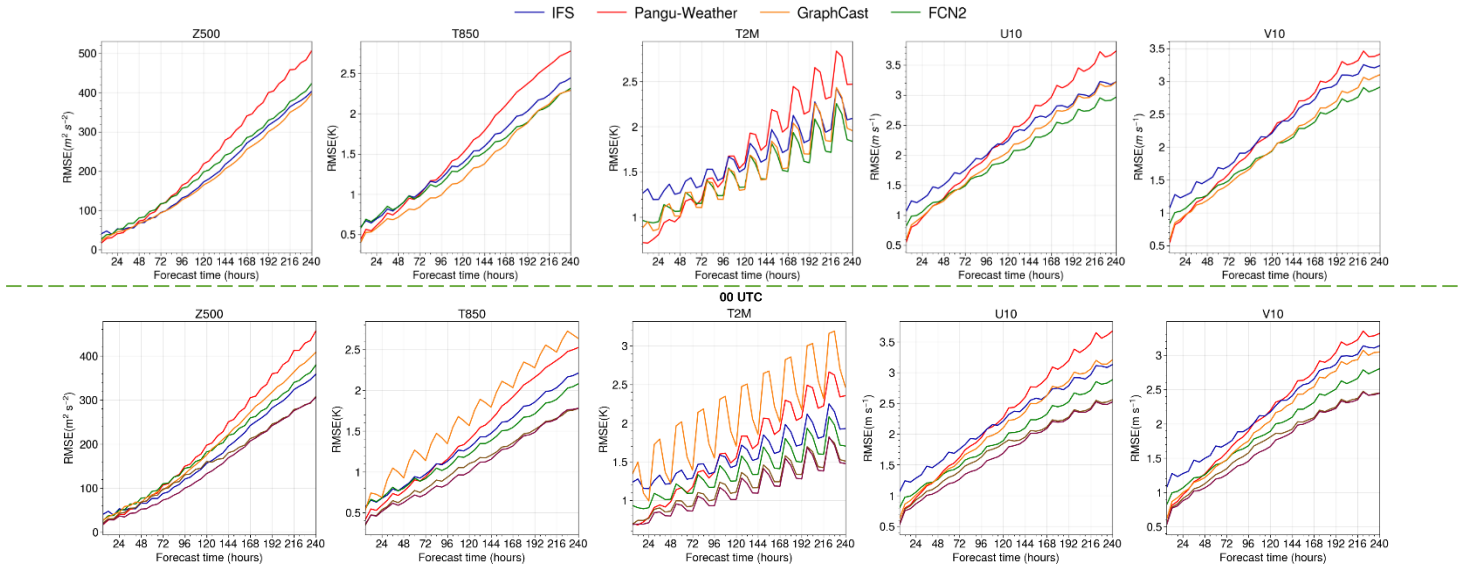
EC

NCEP

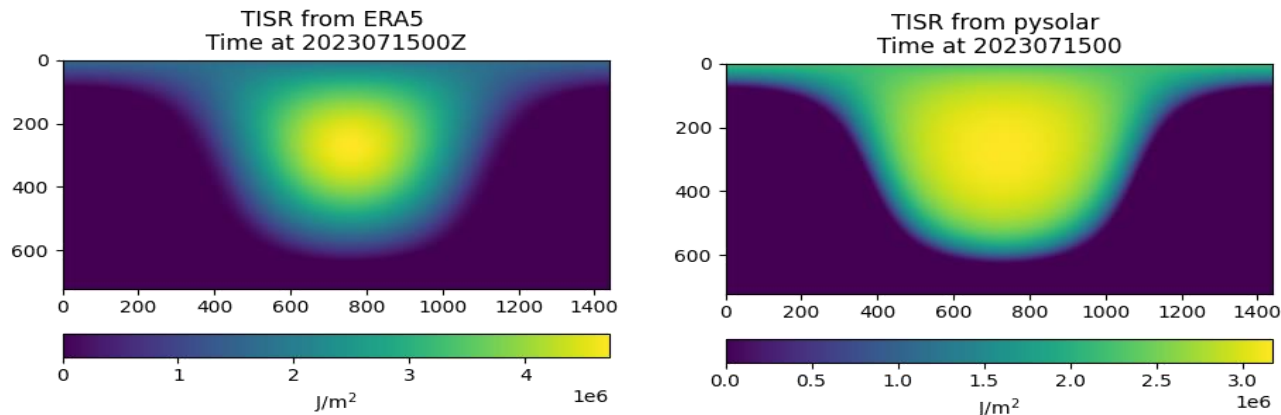
TGFS



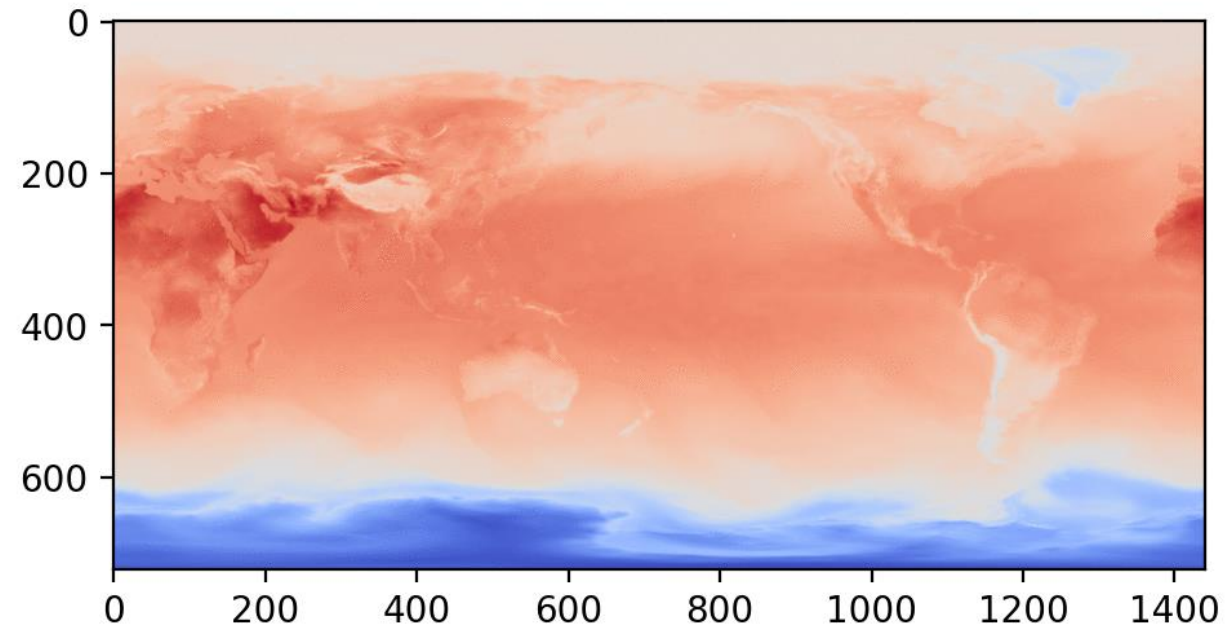
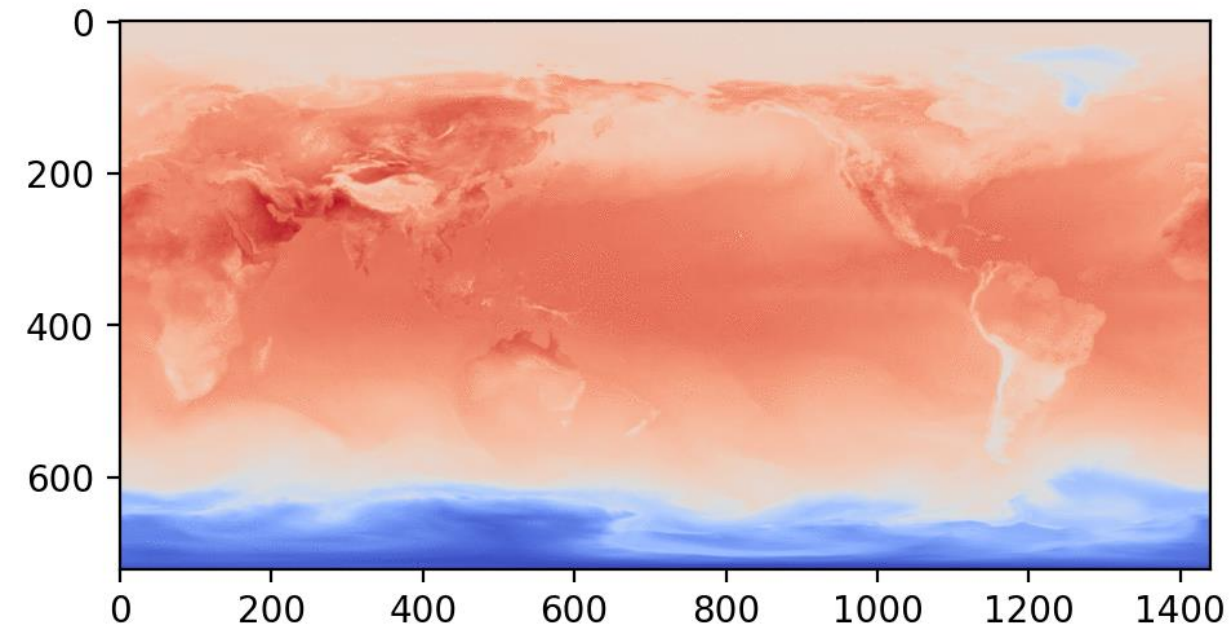
- (大型)全球AI/ML天氣模型之應用經驗
- 有趣的事



上排與下排皆是針對2023年6至8月進行預報校驗，為什麼GraphCast預報表現差這麼多？
 因為GraphCast的輸入有一項Top-of-atmosphere incident solar radiation (tISR)變數，上排的結果是使用ERA5的tISR資料，下排的結果是使用自行估計的tISR資料。因為使用錯的估計套件tISR，導致這個不一致的狀況發生。



- (大型)全球AI/ML天氣模型之應用經驗
- 有趣的事



將輸入的q氣壓層上下顛倒
1000....50 → 50....1000

- (大型)全球AI/ML天氣模型之評估

從應用(快速測試)的經驗，嘗試評估5種模型表現，有必要嗎？

→文章的評估結果皆是在2022年之前，那**2023年之後呢？**評估結果是否如同發表文章結果雷同

→**2023年是歷史上相對熱的一年**，AI/ML應該沒有學過類似的資料，如果評估表現得不錯，增加對AI/ML的信任

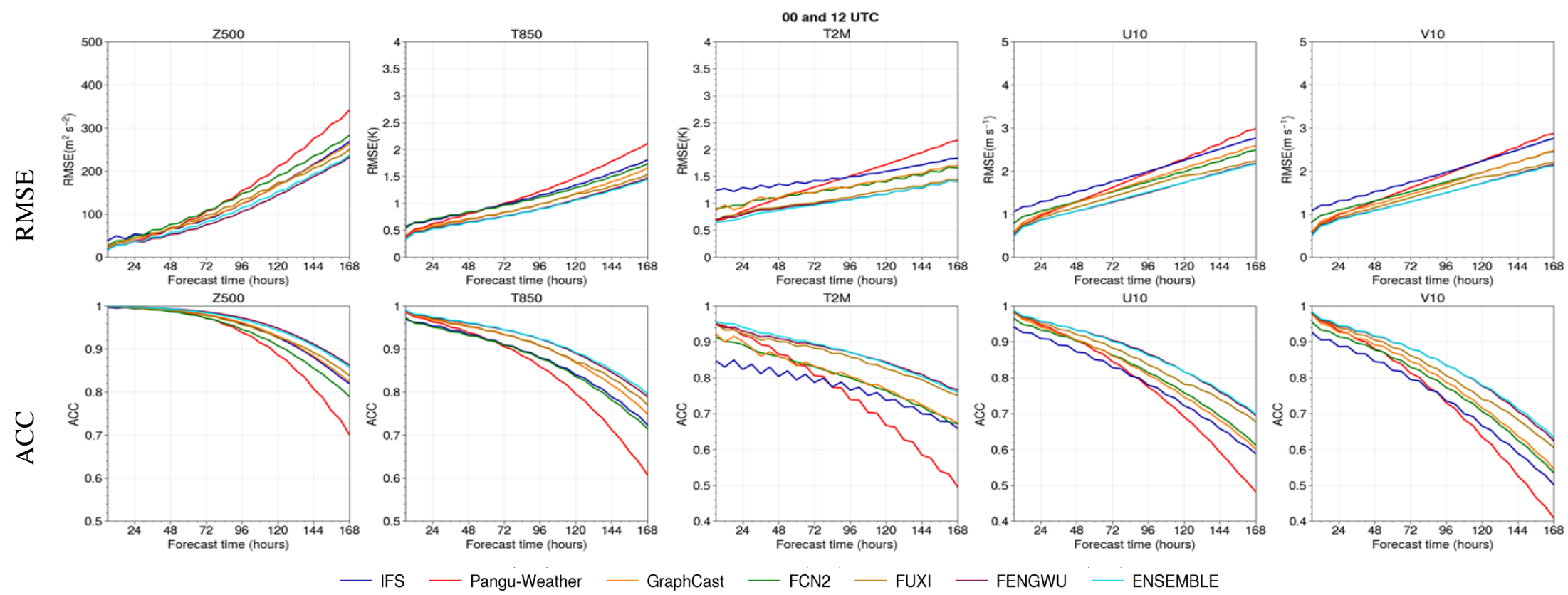
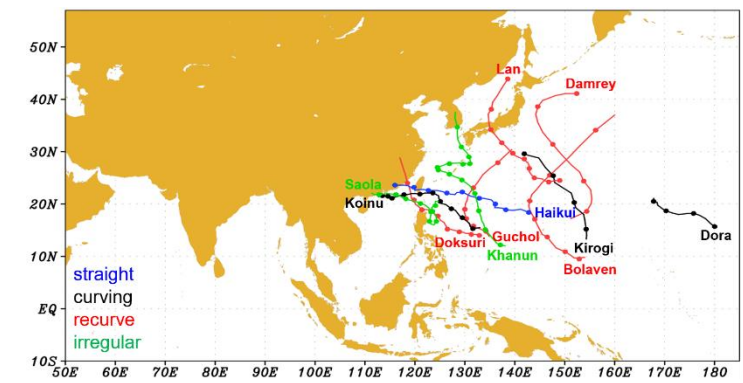
→期待了解AI/ML的能與不能，並該**如何運用「能」**，以及該**如何改善從「不能」到「可能」進而「能」**

• (大型)全球AI/ML天氣模型之評估

五種機器學習全球天氣預報模型 評估-東亞地區

初始場-ERA5、初始時間-00及12 UTC、
評估時間長度-2023/06/01 ~ 2023/11/30

JJASON部分層場校驗結果



— IFS — Pangu-Weather — GraphCast — FCN2 — FUXI — FENGWU — ENSEMBLE
 ENSEMBLE = FENGWU > FUXI > GraphCast = FCN2 > IFS > Pangu-Weather

• 結論與建議

1. 熟悉工具，每一步驟實質確認，常常Github與討論
2. 氣象署是AI應用於天氣的實驗場(資料、軟硬體)
3. 應用端及研發端的並行進而與產業鏈結是重點
4. 關注本國人工智慧AI基本法，知悉應遵守事項
5. 發展高解析度AI/ML天氣模型
6. 全球尺度AI/ML天氣模型預報之應用(區域模式之邊界條件、系集預報應用)
7. 全球尺度AI/ML天氣模型與降尺度之結合



- 結論與建議

AI發展快速，時不時有新的進展，不過世界愈快，心則慢，但腦袋手腳還是要動

接受人工智慧(AI)與機器學習(ML)技術，並確定物理模型與機器學習間的最佳平衡點以持續精進預報



AI並非萬能，擅(善)用AI，不要為了AI而AI

2 5 8 11 14 17 20 23 26 29 32 35 38 41 44 47 50 53 56
Wind Speed (m/s)

0 5 10 15 20 25 30 35
Temperature (°C)

*Thank you
for
your time and
attention.*



**CENTRAL
WEATHER
ADMINISTRATION**