

中央氣象署 115 年第四十屆天氣分析與預報研討會

# 基於 CREDIT 框架之區塊式自回歸多步天氣預報

楊榮瑜 洪語澤 劉正欽 | 交通部中央氣象署 科技發展組

# 研究動機：自回歸步數越多,成本與誤差越高



預報回饋成下一步的輸入 —— 誤差逐步累積

|           | 時間解析度 | 預報長度  | 自回歸步數 |
|-----------|-------|-------|-------|
| ERA5（本研究） | 6 h   | 240 h | 40    |
| CWA區域MLWP | 1 h   | 72 h  | 72    |

## 研究目標:減緩誤差累積影響

# 方法1：一次輸出 4 個時間步,推論步數減少 75%

## 單步輸出



72 h → 模型推論 72 次

## chunk 輸出（4 個時間步）



72 h → 模型推論 18 次

|           | 時間解析度 | 預報長度  | 自回歸步數 | 多輸出步數 |
|-----------|-------|-------|-------|-------|
| ERA5（本研究） | 6 h   | 240 h | 40    | 10    |
| CWA區域MLWP | 1 h   | 72 h  | 72    | 18    |

- 區域逐時示意:72 → 18 次模型推論;本研究 ERA5:40 → 10 次
- 參考影像生成的經驗(MAGI-1、MIMO)，多輸出有助於降低誤差累積的影響
- 理論上來說，少 75% 推論可以減少作業推論時間，也能減少誤差累積

## 方法2：自回歸微調(Autoregression rollout fine-tune)

---

- 現有做法常使用自回歸微調(Autoregression rollout fine-tune)降低誤差累積
- 自回歸微調(Autoregression rollout fine-tune)讓長期預報更穩定,但預報也更平滑
- 但平滑可能刪掉有意義的訊號,也可能刪掉錯誤細節

變平滑是訊號流失,還是錯誤減少？

因此分開討論兩件事 ——

預報保留多少小尺度功率？ 預報的小尺度誤差有多大？

# 實驗

| 編號 | 設定(使用十年ERA5資料訓練,六小時一個時間步) | 模型訓練時間                |
|----|---------------------------|-----------------------|
| A  | 1-in / 1-out 單步預訓練        | 96 node-hours         |
| B  | 4-in / 4-out chunk 預訓練    | 96 node-hours         |
| C  | A 再做 4 步自回歸微調(監督至 24 h)   | 120 node-hours(96+24) |

- 測試集:2020 全年 366 個初始場、10 天預報、40 個時間步
- 模型:wxformer,保持同超參數, 只有學習率退火步數根據epoch修正
- RMSE:全球  $\cos(\text{lat})$  加權,PSD:緯度30–60° 中緯度緯圈平均

# 評估：有多少細節,不等於細節畫得準

## 問的問題

預報功率降到真值一半的波長

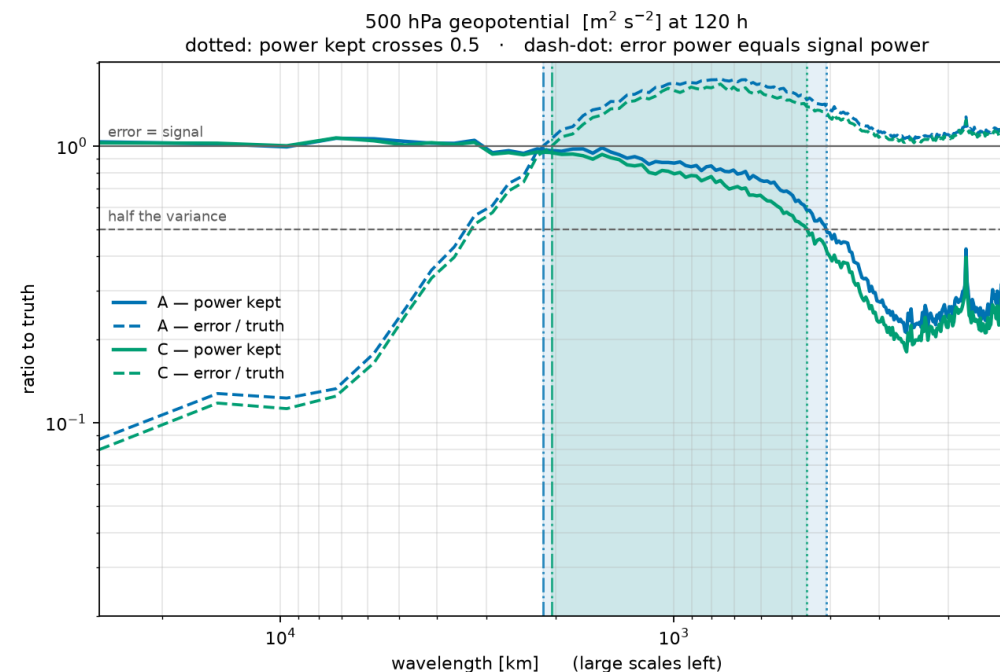
模型保留多少細節？

誤差功率追上真值功率的波長

這些細節錯得多不多？

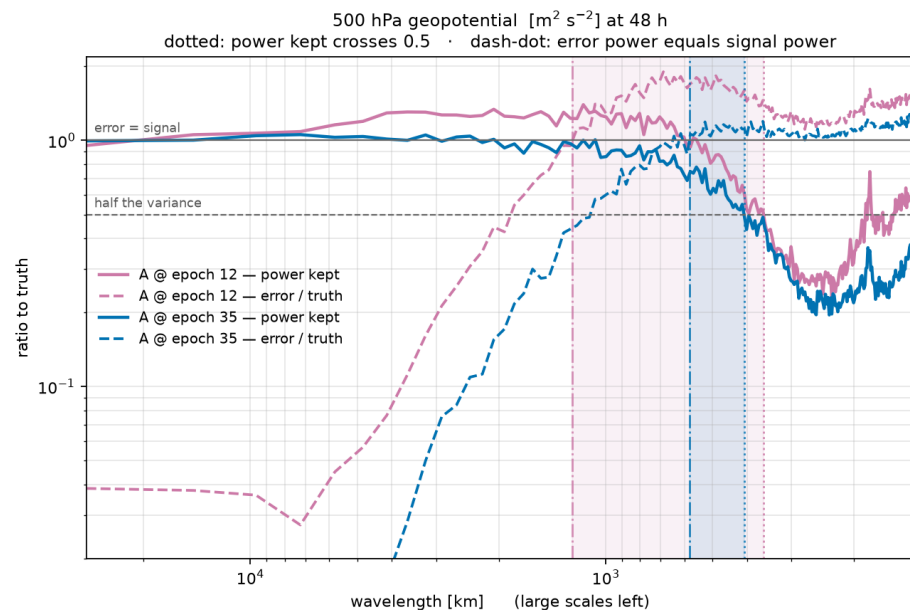
- 預報功率降到真值一半的波長:  $\text{PSD}(\text{預報}) / \text{PSD}(\text{真值}) = 0.5$
- 誤差功率追上真值功率的波長:  $\text{PSD}(\text{預報}-\text{真值}) / \text{PSD}(\text{真值}) = 1.0$

前者看模型還剩多少細節；後者看這些細節錯得多不多。



越往右尺度越大

# 預報功率降到真值一半的波長沒有反映 RMSE 改善



| 同一模型 @48h          | 預報功率降到真值一半的波長   | 誤差功率追上真值功率的波長   | RMSE         |
|--------------------|-----------------|-----------------|--------------|
| 500 hPa 緯向風 ep12   | 541.8 km        | 1,347.5 km      | 4.065        |
| ep35               | <b>544.2 km</b> | <b>753.5 km</b> | <b>2.951</b> |
| 500 hPa 重力位高度 ep12 | 359.6 km        | 1,240.7 km      | 294.6        |
| ep35               | <b>408.1 km</b> | <b>581.9 km</b> | <b>144.3</b> |

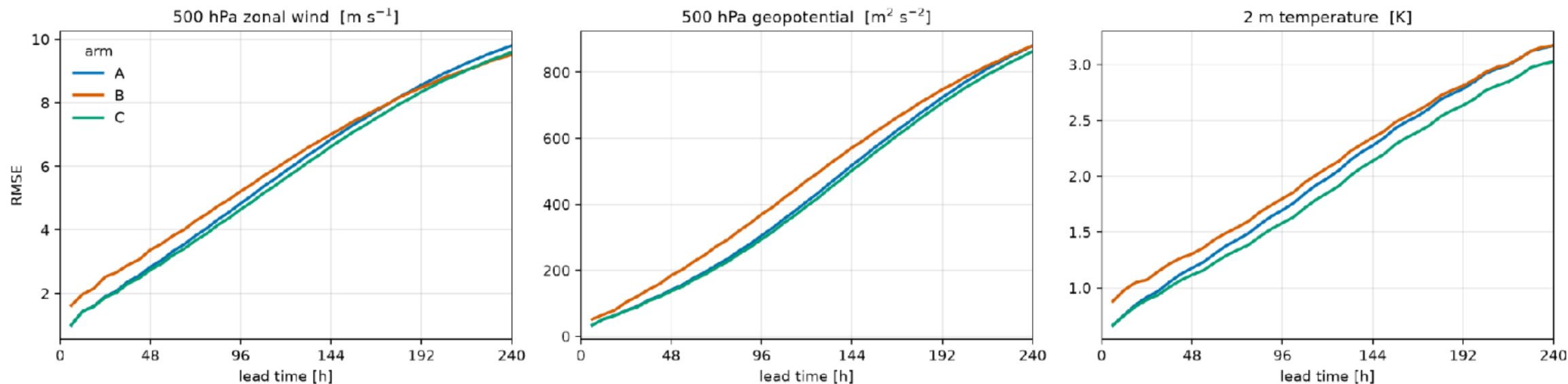
- U500:RMSE 改善 27.4%,預報功率降到真值一半的波長卻幾乎不變 (541.8→544.2 km)
- Z500:RMSE 改善 51.0%,預報功率降到真值一半的波長反而變粗 (359.6→408.1 km)
- 但考慮誤差功率追上真值功率的波長，ep35比ep12都有更好的表現，換言之，雖然ep12跟ep35保留的能量差不多，但ep12在高頻部分可能大多是錯誤的細節

**結論:預報功率降到真值一半的波長不能單獨用來挑選模型版本。**

2020 年 1 月 30 個 初始場;同一模型的 epoch 12 與 epoch 35。

# 結果：B(4-in / 4-out chunk預訓練) 少 75% 推論,但 RMSE 沒有改善

Latitude-weighted RMSE vs lead time (366 init times)

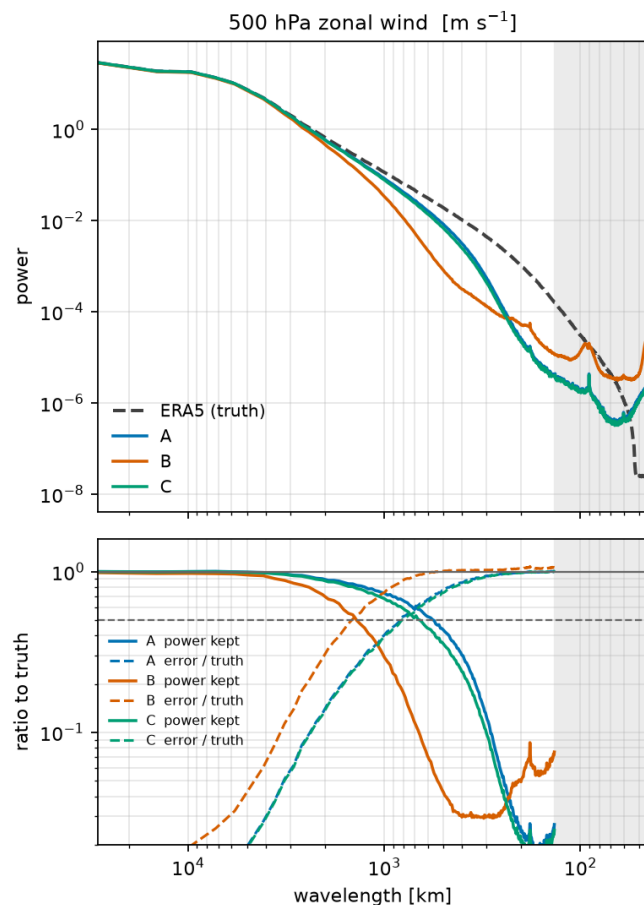


| Z500 RMSE               | 6 h  | 24 h  | 240 h |
|-------------------------|------|-------|-------|
| A 1-in / 1-out 單步預訓練    | 33.6 | 78.3  | 880.4 |
| B 4-in / 4-out chunk預訓練 | 51.7 | 103.2 | 880.4 |
| C A再做 4 步自回歸微調          | 33.9 | 77.2  | 863.5 |

- B:短期預報較差(Z500 6 h +53.9%);到 240 h 與 A 幾乎相同
- C:中長期預報較低(Z500 120 h -3.4%)

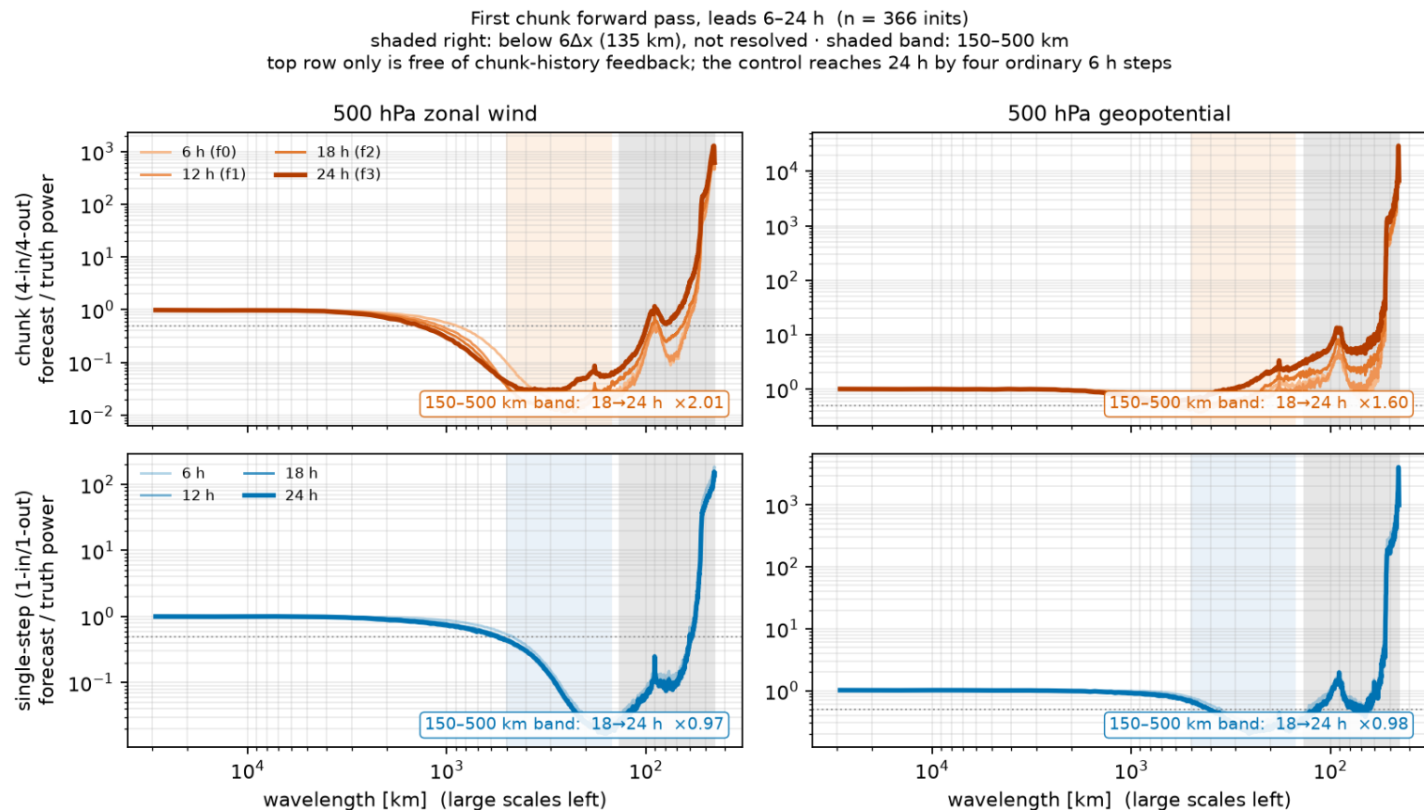
# 結果：B(4-in / 4-out chunk預訓練)在第一次模型推論就流失 U500 中尺度功率

Zonal spectra at 24 h lead ( $|\text{lat}|$  30.0-60.0 [deg], 366 inits)  
solid: power kept (0.5 = half the variance); dashed: error / truth (1.0 = error power equals signal power)  
wavelengths band-representative at  $\bar{\phi} = 43.8^\circ$  (one wavenumber spans  $\times 1.73$  across the band); grey = below  $6\Delta x$ , not resolved



- 24 h U500中，B在1400km左右，功率就掉到實際的一半，A跟C在600km左右才掉下去。
- B 從較大的尺度開始，誤差就與真值訊號同量級
- 24 h 仍在 B 第一次模型推論內，不是多次自身回饋後才出現

# 結果：B(4-in / 4-out chunk預訓練) 的第 4 個輸出突然增加小尺度功率



- 在150–500 km 波段,從18→24 h, B 的 U500  $\times 2.01$ 、Z500  $\times 1.60$
- A 的 U500、Z500 幾乎不變;B 的 t2m 也沒有跳升( $\times 1.04$ )
- 四個時間步來自同一次模型推論, 可以說明不是多次自回歸的問題
- 目前原因未知, 有待後續實驗

# 結果：C 更平滑,但 RMSE 更低

| Z500 @120 h          | RMSE  | 預報功率降到真值一半的波長 | 誤差功率追上真值功率的波長 |
|----------------------|-------|---------------|---------------|
| A 1-in / 1-out 單步預訓練 | 408.2 | 409.8 km      | 2,149 km      |
| C A再做 4 步自回歸微調       | 394.1 | 458.2 km      | 2,042 km      |

- 預報功率降到真值一半的波長:C 458.2 km、A 409.8 km， 可以看出C 較平滑
- 但參考誤差功率追上真值功率的波長:C 2,042 km、A 2,149 km， C雖然較平滑， 但有效細節卻比A多

# 結論

---

## 1. B(4-in / 4-out chunk 預訓練) 省 75% 推論,但 RMSE 沒有改善。

A/B 訓練預算相同;B 將十天預報由 40 次模型推論降至 10 次,但短期預報較差,到 240 h時才與 A 幾乎相同。

## 2. 預報功率降到真值一半的波長不能單獨判斷預報好壞。

根據A實驗不同epoch模型的比較, 35個epoch的U500 的 RMSE 改善 27.4%,預報功率降到真值一半的波長幾乎不變;

Z500 的 RMSE 改善 51.0%,預報功率降到真值一半的波長反而由 359.6 變成 408.1 km。

但誤差功率追上真值功率的波長都是35個epoch的模型更好

## 3. 更平滑不一定更差。

模型在預報能量保留比較好, 不代表保留的是有效資訊, 所以評估上除了要看預報的能量, 也要看誤差的頻譜分析

## 4. B(4-in / 4-out chunk 預訓練)在同epoch下表現比A好 我們推測是欠擬合

未來工作會進行增加B的訓練次數的實驗, 並將有效的研究成果遷移到區域模型的訓練

# 謝謝聆聽

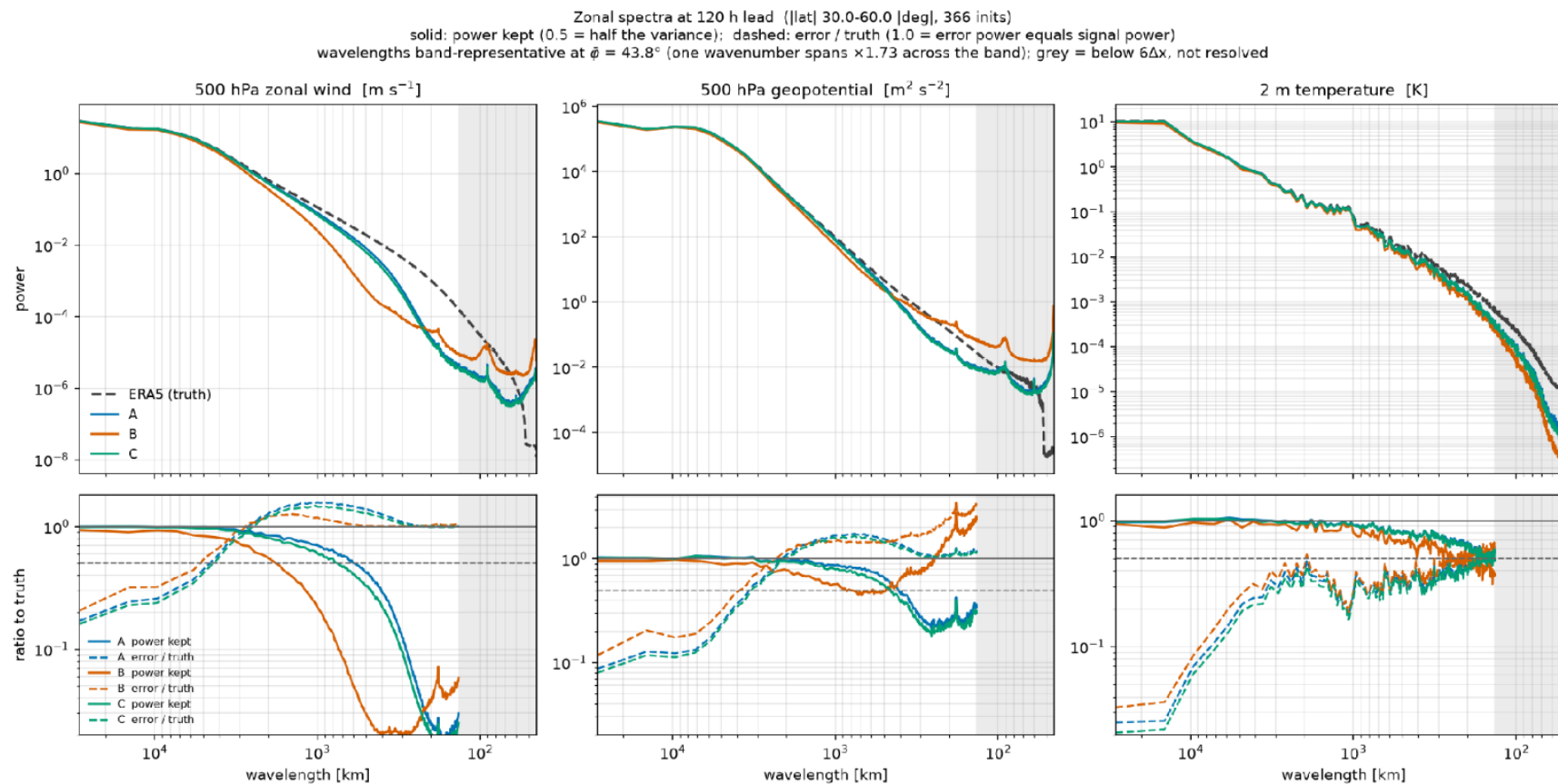
楊榮瑜 洪語澤 劉正欽 | 交通部中央氣象署 科技發展組

# 備用：這個結論能外推到哪裡

|                        | ERA5（本研究）                       | 區域 2 km / 1 h / 72 h | 外推狀態              |
|------------------------|---------------------------------|----------------------|-------------------|
| 模型推論數（各自所列預報長度）        |                                 | 40 → 10              | 72 → 18 相同：皆減 75% |
| 絕對少做的推論數               |                                 | 30                   | 54 不同             |
| 最遠直接預報時間步              |                                 | +24 h                | +4 h 不同           |
| 單幀任務難度                 | —                               | —                    | 未知                |
| 實際端到端加速比               | —                               | —                    | 未知                |
| 逐時側邊界 / future forcing | 不適用                             | —                    | 未知                |
| 幀 3 功率異常是否重現           | U500 ×2.01、Z500 ×1.60、t2m ×1.04 | —                    | 未知                |

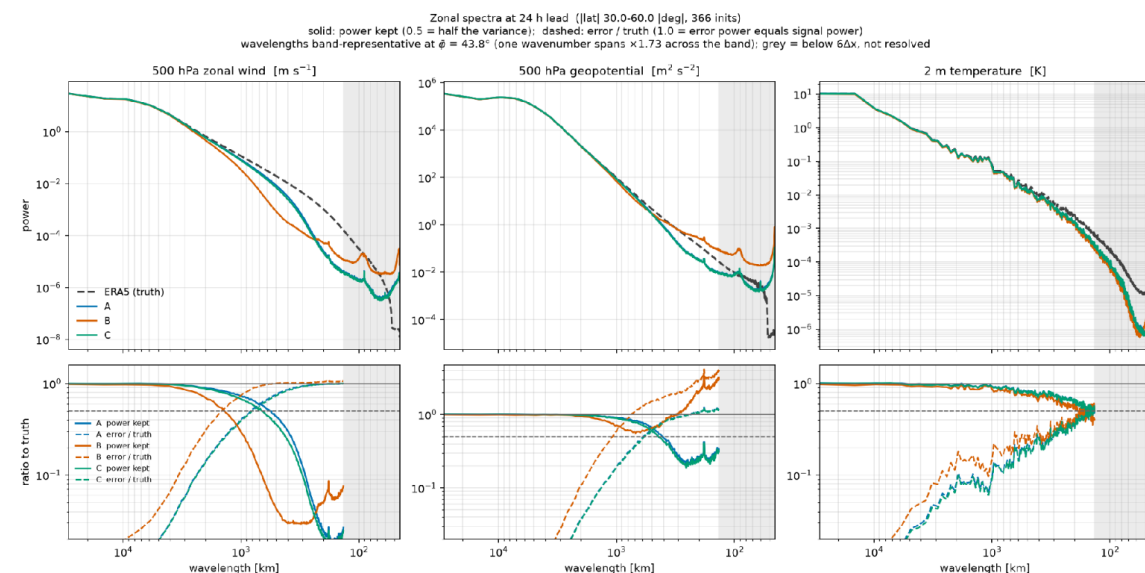
本研究的負結果限於:6 小時間步長、24 小時 chunk 跨度、單一 seed。  
「區域設定收益較大」需要在該設定實測後才能宣稱;本頁只列出可查證的差異。

# 備用：lead 120 h 功率譜



- B (橘) 在中尺度的功率保留仍較粗,方向與 24 h 一致
- 灰區為  $6\Delta x$  以下,兩條 ratio 曲線皆已排除

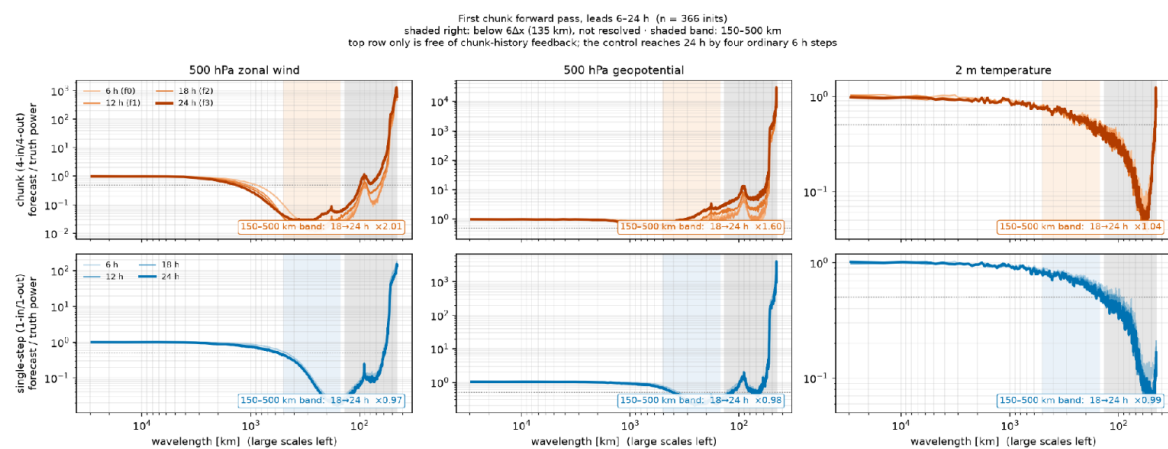
# 備用：三變數完整版



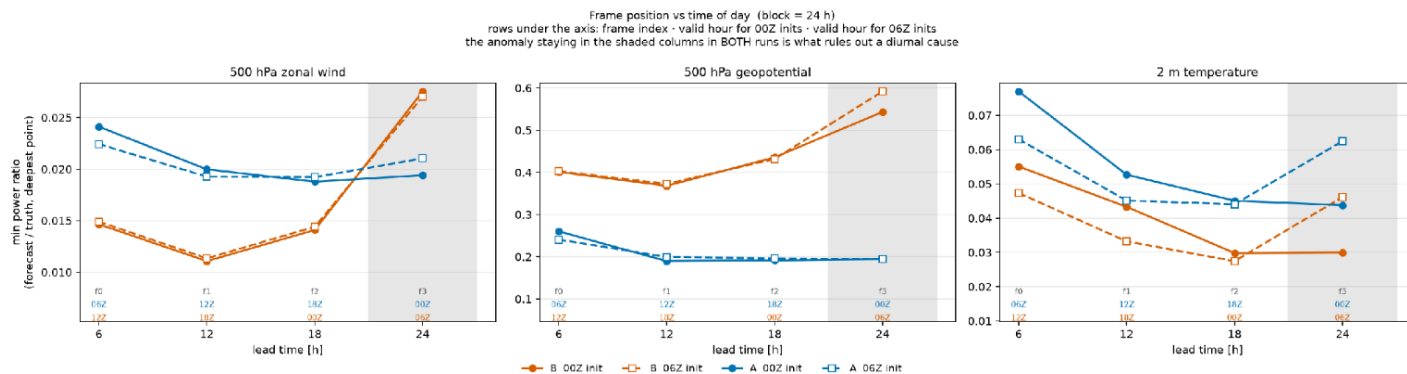
上:lead 24 h 功率譜與兩條 ratio 曲線

下:第一個 chunk forward pass 的完整比值曲線

- t2m 的誤差=訊號截止落在  $6\Delta x$ (135 km)以下,在正文中不作物理解讀
- t2m 在 150–500 km 的幀 3 變化為  $\times 1.04$ ,與 500 hPa 兩場不同
- 分兩頁而非六格並排,是因為六格在會場後排看不清



# 備用：06Z 初始場對照



- block = 24 h,所以在 00Z 初始場下,幀位置與一天中的時刻完全綁在一起
- 改 06Z 初始場:幀 3 由 00Z 移到 06Z,時間步仍是 24 / 48 h
- 尖峰留在時間步 24 / 48 h(灰底欄)→ 排除「固定 UTC 時刻」
- 單步對照兩條皆平
- 只排除這一個替代解釋;輸出槽、直接時距、loss distance、forcing age 仍共線
- $n = 30$ (2020 年 1 月),與正文的 366 例不同

# 備用： 幀 3 異常尚未隔離的替代機制

| 幀  | 相對 block 起點 | forcing age | 日變相位差          | 實測功率異常     |
|----|-------------|-------------|----------------|------------|
| f0 |             | +6 h        | 6 h            | 90° —      |
| f1 |             | +12 h       | 12 h 180° (最大) | —          |
| f2 |             | +18 h       | 18 h           | 270° —     |
| f3 |             | +24 h       | 24 h 0° (同相)   | U500 2.29× |

- 一次 forward 內四幀共用 block 起點的 forcing,故 forcing age 與輸出槽完全共線,改初始場時刻無法拆開
- 若成因為 forcing 相位錯誤,最差的應是 f1(相位差 180°),而非同相的 f3
- t2m 對日照 forcing 最敏感,卻是唯一無幀效應的變數(1.02×)
- 但以上兩點皆以功率指標得到,功率指標檢驗不了相位假說 —— 列為未排除,不作為反證
- 可拆解的設計:逐時間步 future-forcing 對照;四個 4-in/1-out 的 horizon-matched 對照