

台灣區域機器學習天氣預報模式之 設計與初步驗證

Design and Preliminary Verification of a
Regional Machine-Learning Weather Prediction Model for Taiwan

洪語澤 楊榮瑜 劉正欽 連國淵

中央氣象署 科技發展組

1 Why a regional MLWP model?

1 Inference cost

Second-scale inference makes large ensembles and high-frequency updates practical — out of reach for physics-based NWP at this resolution.

2 Resolution gap

Global MLWP models are already skilful, but a 25–31 km grid cannot resolve Taiwan's convection.

3 Terrain-scale phenomena

Taiwan's high-impact weather, from orographic rainfall to typhoon circulation, is governed by terrain.

Training framework: CREDIT / WXFormer

CREDIT

Community Research Earth Digital Intelligence Twin — MILES group, NSF NCAR

- End-to-end pipeline: preprocessing → training → inference → verification
- Modular — supports multiple architectures (FuXi, WXFormer) .
- Scales on standard HPC systems

WXFormer

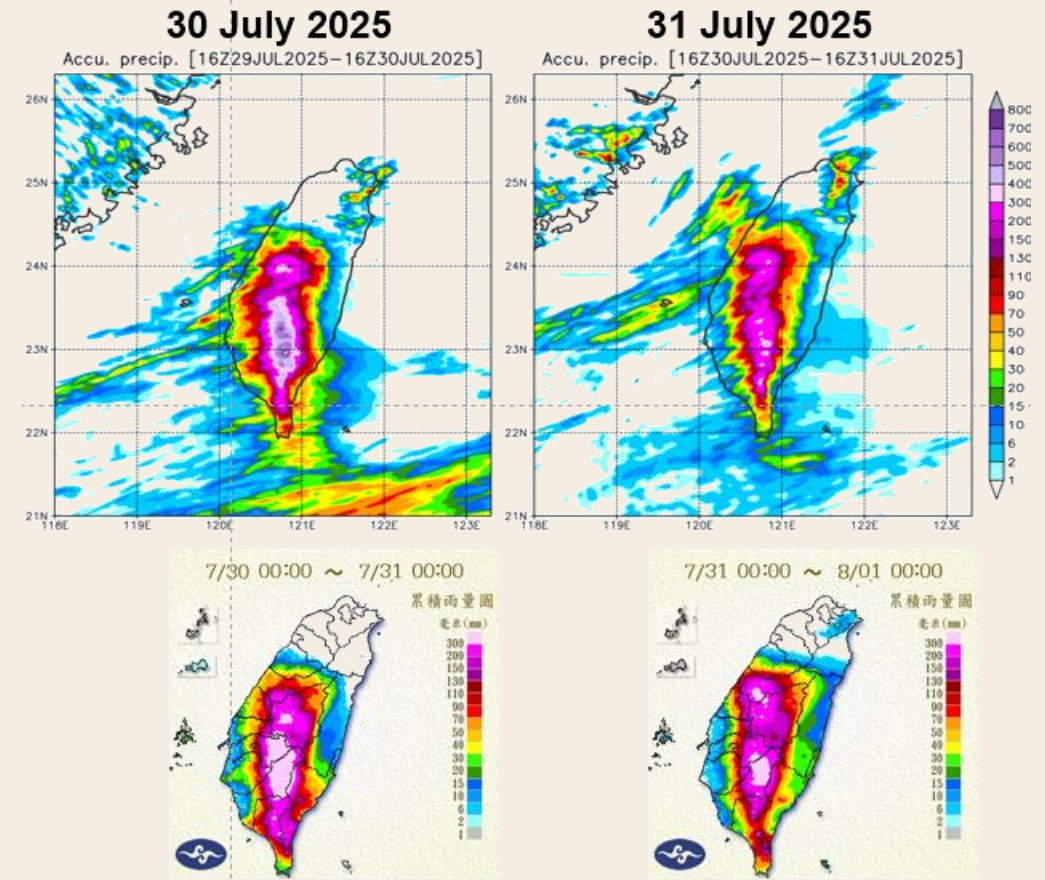
Deterministic multiscale vision transformer, autoregressive

- CrossFormer encoder + hierarchical transposed-convolution decoder with skip connections
- Spectral normalisation, intelligent padding and multi-step training to curb compounding error growth

TARRA: Taiwan Regional Reanalysis

CWA TAIwan Regional ReAnalysis (tentative name)

Base system	CWA RWRF (WRF + WRFDA)
Resolution	2 km, 450*450
Assimilation	Hourly radar DA (reflectivity, radial wind) + surface DA
Output frequency	Hourly analyses
Period	1 year (2025) ; 10 years (2017–2026) expected end of 2026
Large scale	Blending and boundary conditions from ERA5



4

Constraint: domain size and predictability

900 km

450 × 450 at 2 km

TARRA domain covers the vicinity of Taiwan

12 hours

at a steering flow of 20 m/s

time for a parcel to cross the domain

4 times

over a 48-hour forecast

the interior is flushed by the boundary

Within the target lead time, the synoptic-scale solution is set by the boundary conditions

What this constraint means

- Not "the model is useless" — rather, its value does not lie at the synoptic scale
- Nor can it be solved by enlarging the domain — the domain is bounded by the training data
- The question becomes: if BC sets the synoptic scale, where does the skill come from?

5

Terrain: a source of predictability, and a threat

× 10

difference in error-growth rate, and in saturation time, between convective and synoptic scales

1.5 h / 54 h

where the tangent-linear approximation breaks down — cloud-resolving vs synoptic

~ 4 h

mean error-doubling time reported for limited-area cloud-resolving models

Hohenegger & Schär (2007), Bull. Amer. Meteor. Soc. 88, 1783–1793 and J. Atmos. Sci. 64, 4467–4478; Zhang, Snyder & Rotunno (2003), J. Atmos. Sci.

Deterministic convective-scale skill is therefore gone long before 48 h — what survives is the terrain-locked response

Opportunity

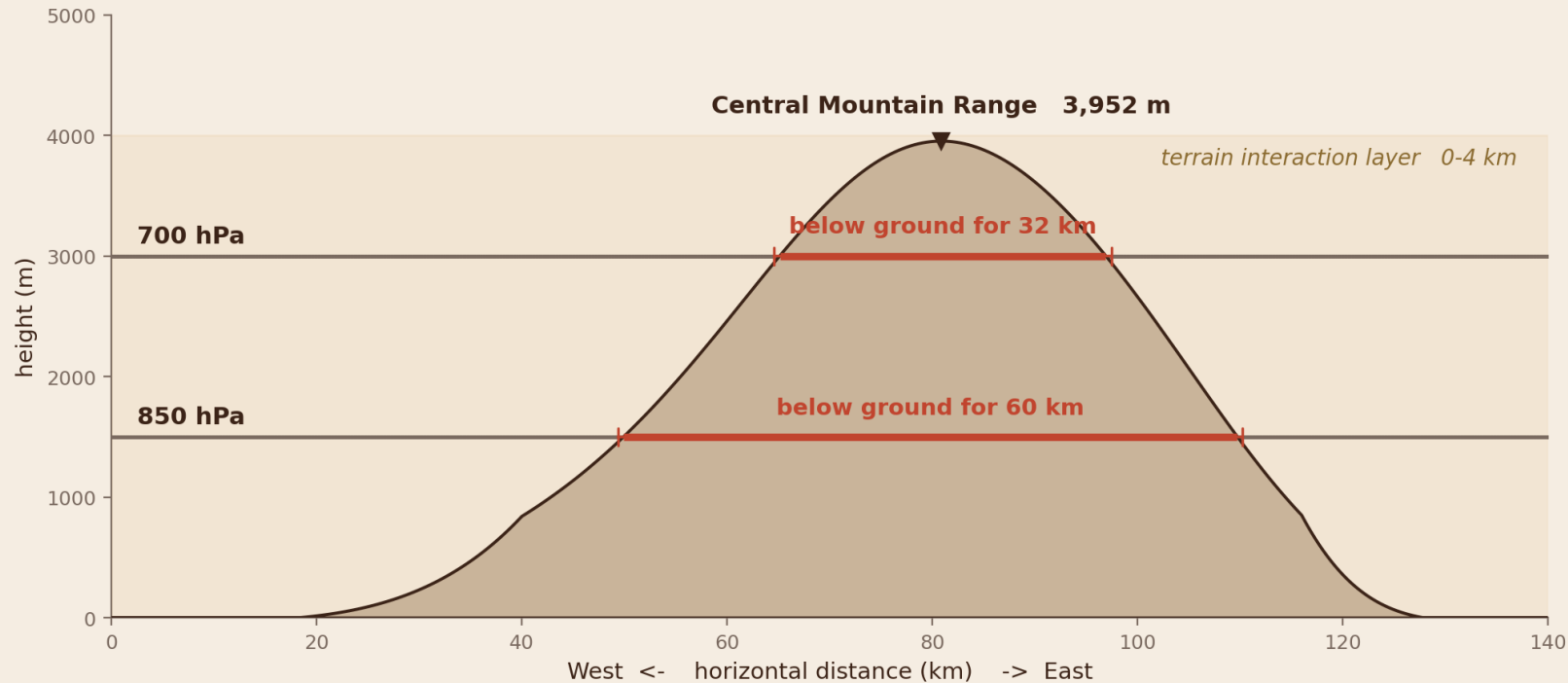
The orographic response does not randomise. Upslope lifting, lee vortices and rainfall hot spots recur in the same places whenever the inflow is right — so a 2 km model can still add value at the target lead time.

Threat

The same terrain breaks the standard way of representing the data over Taiwan. Handled badly, we lose the benefit entirely and add systematic error on top.

6

Terrain challenge: pressure levels break down



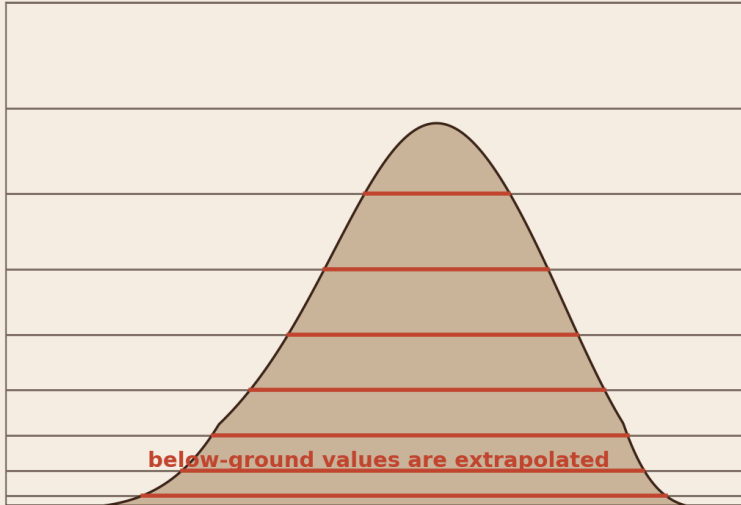
The problem

- 850 hPa ($\approx 1,500$ m) and 700 hPa ($\approx 3,000$ m) lie below ground over much of it
- Global models and reanalyses give downward-extrapolated, non-physical values there
- And 0–4 km is exactly where all terrain interaction happens

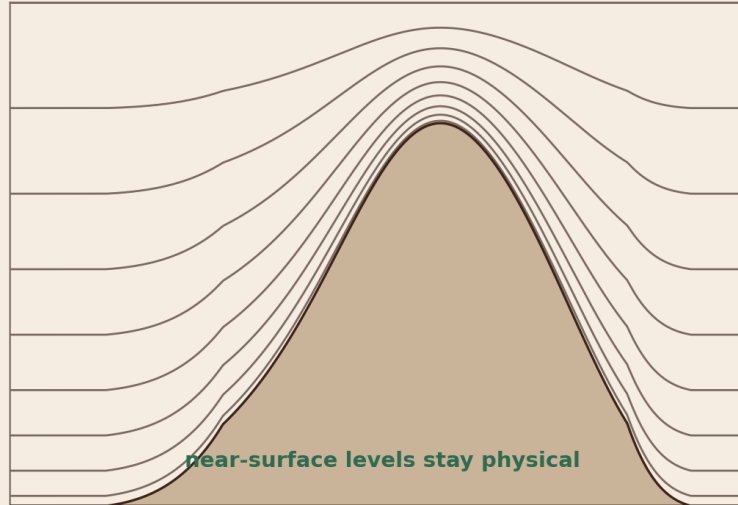
Training on pressure levels = learning from fabricated data where orographic rainfall matters most

Vertical coordinate: eta levels

Pressure coordinate: levels cut into terrain



Eta coordinate: levels follow terrain



Rationale

- No below-ground extrapolation; near-surface levels stay physical over the mountains
- Published work reports improved orographic precipitation (Sha et al, 2025)
- TARRA is natively on WRF eta levels (51), so no extra interpolation is needed

Requirement: the model must know the coordinate itself

- The same eta level maps to very different heights and pressures over ocean and over peaks
- Terrain height (static) and surface pressure (prognostic) must be model inputs, so each level's actual state can be recovered

Only use 15 layers currently. More experiments is going on.

8 Why the BC covers the whole domain, not just a rim

Design

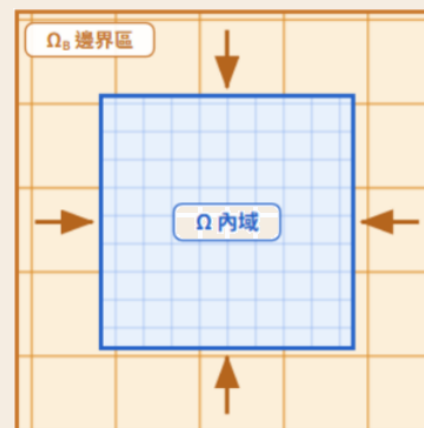
- BC source: interpolated to ~3km ERA5 at the matching lead time
- Updated every hour, at every autoregressive step
- concatenated as extra channels across the whole domain, not only a boundary ring

Why an ML model needs more than a rim

- No governing equations — inward propagation of boundary information has to be learned, not inherited
- No conservation constraints — the interior can drift in ways a dynamical core cannot
- The interior is boundary-controlled anyway; full-domain input makes explicit what advection would deliver

傳統邊界條件（如 WRF NWP）

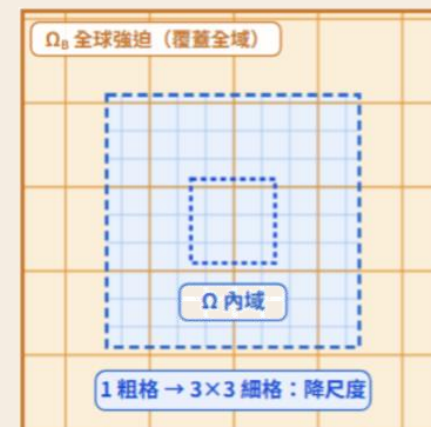
$$\Omega \cap \Omega_B = \emptyset \quad \text{內域與邊界互斥}$$



大尺度資訊僅由外圍邊框（邊界區）提供，
再逐步向內域傳入

重疊域邊界強迫（本研究）

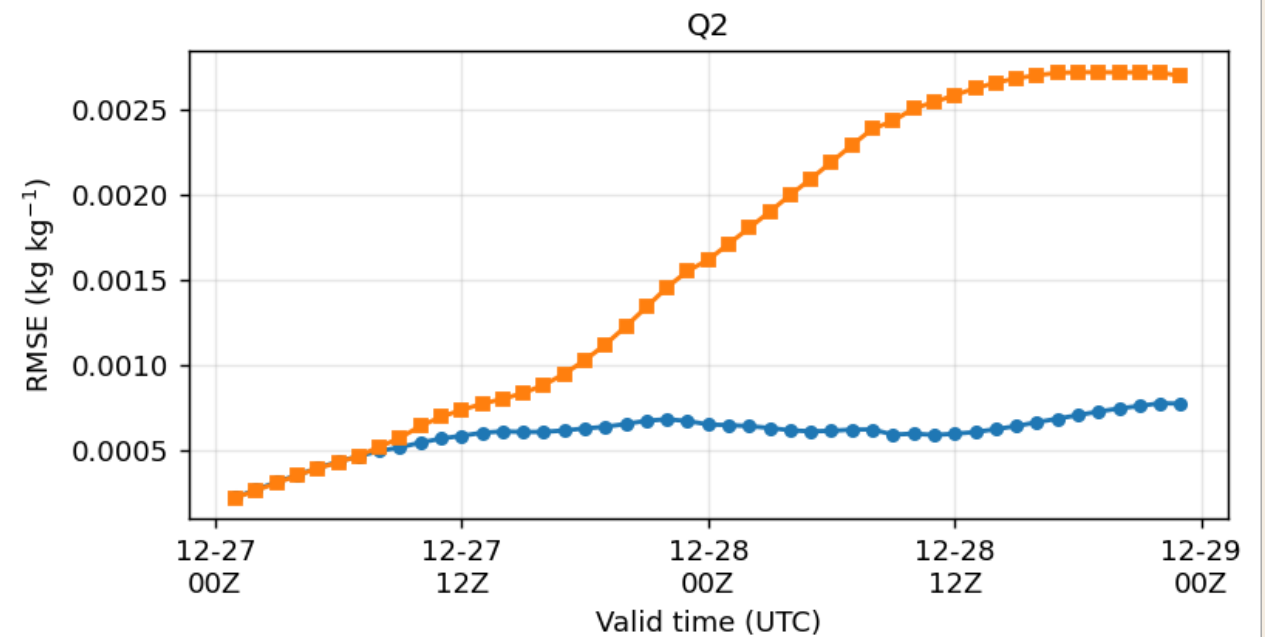
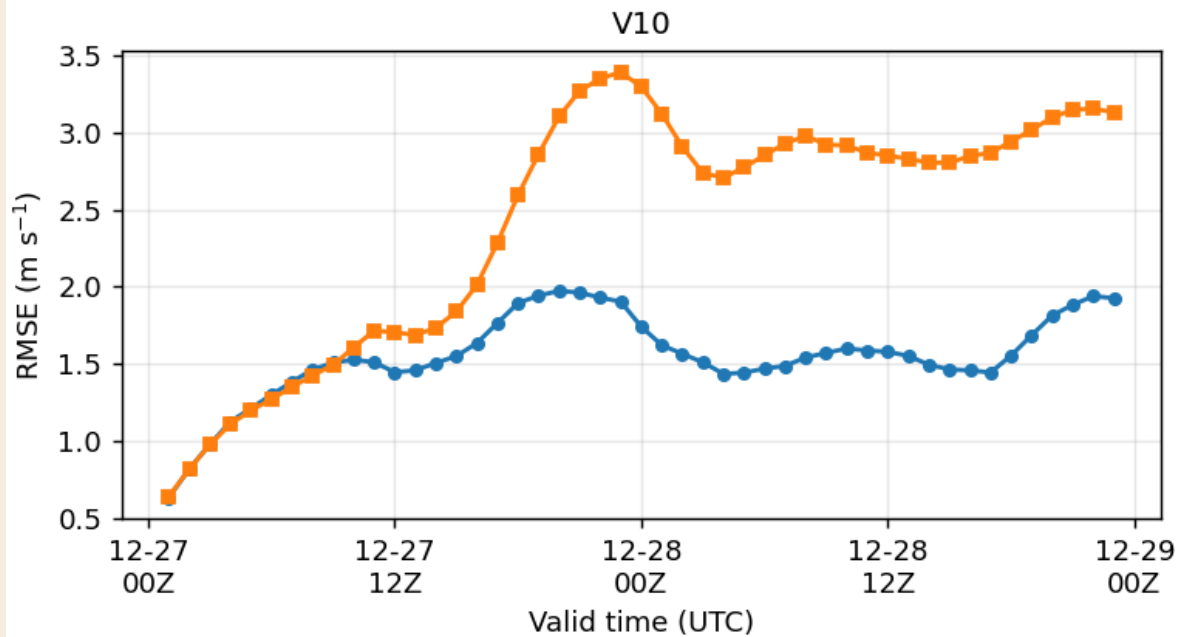
$$\Omega \subset \Omega_B \quad \text{邊界強迫涵蓋整個內域}$$



粗解析全球強迫覆蓋整個內域，模式同時
進行物理積分與降尺度 (downscaling)

Conceptually this is closer to spectral nudging than to lateral boundary forcing

Preliminary result: BC is necessary



What this supports

- BC information is necessary to maintain forecast quality
- Without BC, the model loses the synoptic field rapidly with lead time
- Empirical confirmation that at this domain size the boundary sets the synoptic solution

What it does not support

- The control here is BC vs no BC, not one injection scheme against another
- It cannot establish that full-domain concatenation is optimal
- A single case is not enough to claim overall quality

The risk: the train–inference gap

Layer 1 — distribution shift in the BC source

- Training BC is ERA5; inference BC is a global MLWP forecast
- L2-trained models smooth systematically with lead time, losing high-wavenumber power
- This is distribution shift, not random error → adding noise to the BC is the wrong fix

Layer 2 — TARRA's large scale comes from ERA5

- In training, BC and the target share the same large-scale source
- The model may learn too tight a mapping: ERA5 large scale → TARRA detail
- Full-domain concatenation amplifies this — BC is available at every grid point

Mitigation

Spectral degradation augmentation

Measure the global model's spectral loss by lead time, then low-pass filter the BC accordingly during training. Cheap, and in the right direction.

BC lead time as an input

Lets the model learn that the older the BC is, the more it should rely on its own dynamics.

Multi-source BC

Avoids overfitting to one global model, and naturally supports BC-driven ensembles.

Taiwan-specific risk: global MLWP models underestimate typhoon intensity — smoothing hits exactly where sharpness matters most

Comparison to downscaling

What is shared

- The synoptic-scale solution is effectively set by the global model
- Added value concentrates in the terrain-scale response

What differs

- Downscaling is an instantaneous mapping; this model has its own temporal evolution and internal dynamics
- Error growth and drift exist → long rollouts are required
- It can generate mesoscale systems absent from the global model

Positioned between the two → the baseline must be "global MLWP interpolated to 2 km"

12 Conclusions and outlook

- 1 TARRA supplies the high-resolution, temporally continuous training data a Taiwan regional MLWP model needs
- 2 Domain size makes predictability BC-dominated within the target lead time; our first experiment confirms BC is necessary
- 3 Terrain-locked signal is the primary source of predictability → eta coordinate, vertical budget over the mountain height range
- 4 The main risk is the train–inference gap arising from the global model's spectral loss

Next steps

Vertical level configuration

BC injection comparison

Rainfall (hydrometer/w...)

Typhoon case evaluation

Thank you