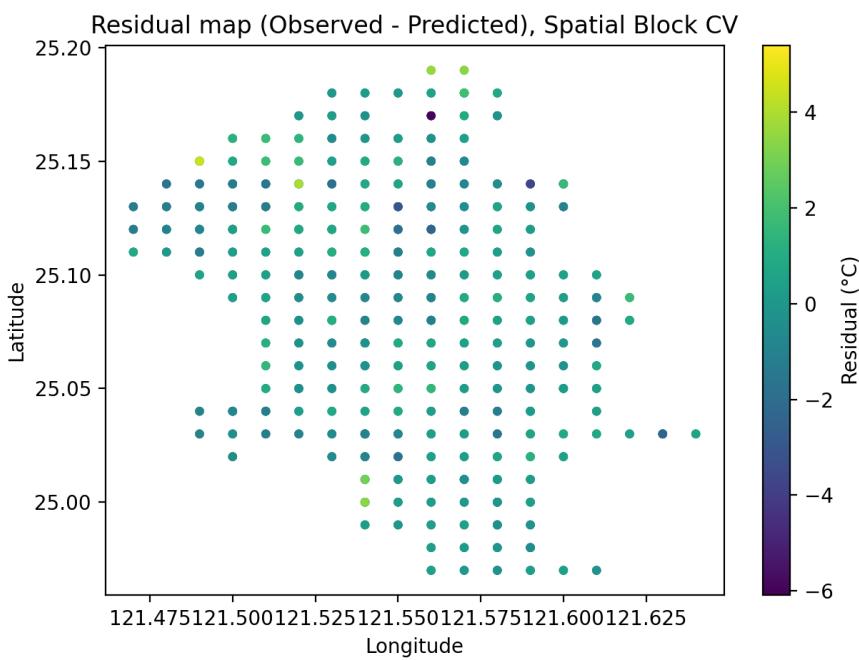


Mini ML Project: Explainable Heat Exposure Prediction

Random Forest • Spatial Block Cross-Validation (anti-leakage) • Permutation importance

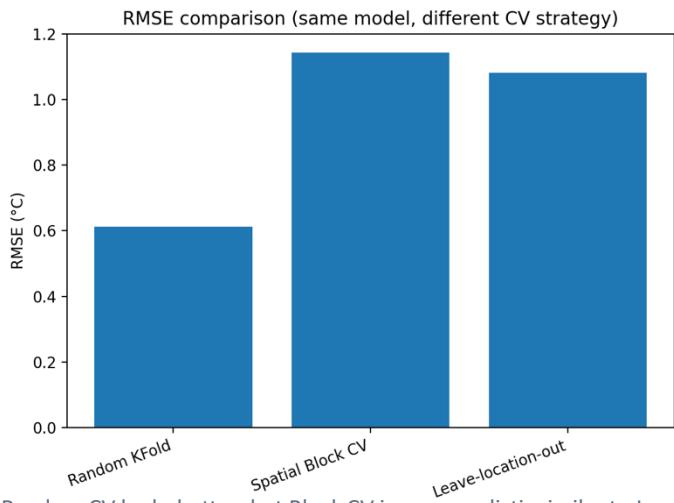
Study area: Taipei (grid points) | Target: mean_t ($^{\circ}$ C) | N = 1104

Data: Figshare DOI 10.6084/m9.figshare.24922275



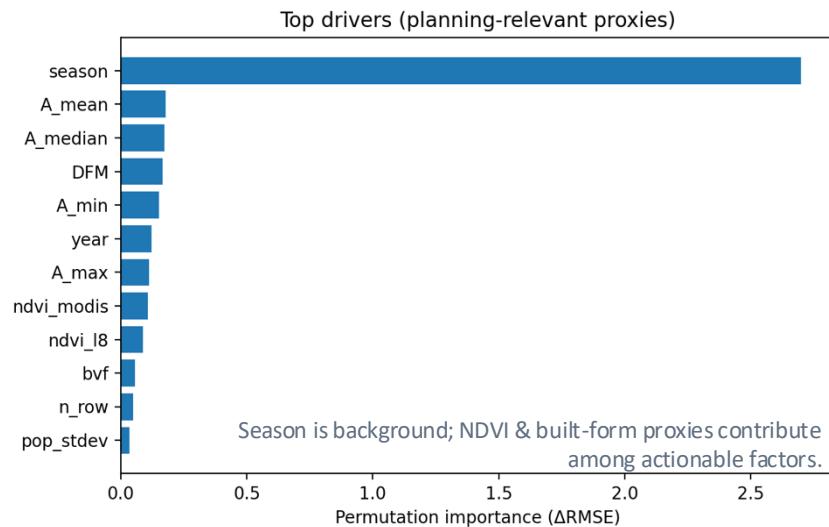
Positive = under-prediction (hotter than predicted); negative = over-prediction.

A. Residual map (Observed – Predicted), Spatial Block CV



Random CV looks better, but Block CV is more realistic similar to Leave Location out model.

B. RMSE by validation strategy



C. Top drivers (Permutation importance, Δ RMSE)

Purpose & setup

- Predict mean temperature from planning-relevant proxies (greenness, morphology, land use, activity).
- Use a non-linear tabular baseline and keep the pipeline explainable.

Validation (avoid spatial leakage)

- Random KFold can be over-optimistic when nearby locations are similar.
- Spatial Block CV holds out entire spatial blocks (new areas).
- Leave-location-out checks repeated site effects.

Core result (RMSE, $^{\circ}$ C)

- Random KFold: 0.61
- Spatial Block CV: 1.14
- Leave-location-out: 1.08
- Drop under spatial CV → confirms leakage risk in random splits.

Explainability

- Permutation importance (Δ RMSE) highlights model reliance.
- Season dominates (background); NDVI & built-form proxies (e.g., BVF) matter among actionable factors.

Key limitations

- Importance is not causality.
- Proxy/scale mismatch: grid temperature vs street-level experience.
- External validity: other cities/years may require re-training.