

Urban Heat Islands and Local Climate Downscaling

GenHack 2025

Ayoub Oulad Ali Maria Cristina Centorame Yuxin Li

Team #23

Abstract

Urban Heat Islands (UHI) emerge from the interaction between land cover, vegetation scarcity, atmospheric circulation, and local microclimates. In the GenHack 2025 challenge, we investigated how reanalysis products (ERA5 and ERA5-Land), vegetation indices from Sentinel-3, and ground-station observations jointly describe these dynamics.

Across four periods of work—data exploration, visualization, metric design, and modelling—we constructed a unified pipeline for quantifying ERA5 biases and learning physically informed corrections at the station level. Our exploratory analysis revealed strong spatial gradients in temperature across Europe, robust seasonal wind patterns, and a clear negative relationship between vegetation (NDVI) and temperature anomalies around Paris.

Building on these insights, we developed a set of corrective residual-learning models, culminating in a Random Forest and a Graph Neural Network capable of removing most of the structured bias present in ERA5. The final models achieve substantial improvements in RMSE while remaining interpretable through feature importance and spatial diagnostics.

Overall, our analysis highlights the joint role of vegetation, altitude, and large-scale atmospheric flow in shaping UHI intensity and demonstrates how multimodal datasets can be combined to improve local climate downscaling.

1 Introduction

Urban Heat Islands (UHIs) are localized temperature anomalies driven by reduced vegetation, impervious surfaces, and the geometry of built environments. These effects amplify heat exposure during warm periods and interact with broader synoptic conditions such as atmospheric stagnation and seasonal wind regimes. Understanding UHI intensity therefore requires combining heterogeneous data sources that capture complementary physical processes across multiple scales.

In the GenHack 2025 challenge, our goal was to analyse and correct the discrepancies between coarse-resolution ERA5 reanalysis temperatures and ground-station observations, using vegetation indices and atmospheric circulation as key explanatory variables. Our work integrated three main types of information:

- **ERA5 / ERA5-Land** reanalysis data, describing large-scale temperature patterns and 10 m wind fields;
- **Sentinel-3 NDVI** imagery, providing vegetation intensity and seasonality around urban areas;
- **Ground-station records** from ECA&D, offering precise point-level temperature measurements.

Over the four periods of the challenge, we progressively explored, visualised, quantified, and modelled the interactions between these datasets. The project evolved from large-scale climatological exploration (Period 1–2) to bias quantification (Period 3) and finally to explanatory residual modelling (Period 4). The resulting framework allows us to correct ERA5 temperatures at station locations using physically interpretable features related to vegetation, altitude, and atmospheric circulation.

This report presents the full pipeline, from exploratory visualisation and dataset fusion to quantitative metric design and the development of a graph-aware residual-learning model capable of improving local temperature estimates while revealing the main drivers of UHI intensity.

2 Data Sources and Preprocessing

2.1 ERA5 and ERA5-Land Reanalysis Data

ERA5 and ERA5-Land provide hourly reanalysis fields at approximately 0.25° resolution. For our purposes, we extracted:

- daily maximum 2 m air temperature (**tmax**),
- 10 m wind components (**u10**, **v10**),

and visualised their spatial patterns across Europe. ERA5 offers consistent synoptic-scale information but tends to smooth extremes and can deviate from observations in complex terrain, coastal zones, or densely urbanised regions.

2.2 Sentinel-3 NDVI Imagery

Vegetation was characterised using Sentinel-3 NDVI GeoTIFFs. The images were:

- cropped spatially to the Paris region,
- rescaled to a common grid for consistent temporal comparison,
- averaged over time to generate a multi-year NDVI time series.

NDVI provides crucial information on vegetation density, seasonality, and interannual variability, all of which influence local cooling capacity and UHI intensity.

2.3 Ground-Station Temperature Observations

Ground-station data were obtained from the ECA&D archive:

- only stations with country code FR were retained, resulting in **44 stations**;
- daily maximum temperature (**TX**) was used;
- standard quality-control filters were applied to remove missing or inconsistent records.

Stations represent the most accurate source of temperature measurements and serve as the reference against which all models are evaluated.

2.4 Unified Station-Level Dataset

To enable bias analysis and residual modelling, we constructed a fused station-level dataset by aligning:

- station temperatures,
- ERA5/ERA5-Land values interpolated at station coordinates,
- local NDVI sampled at station latitude/longitude.

A dedicated routine (`create_snapshot`) ensured spatial and temporal consistency, allowing us to extract daily or seasonal “snapshots” combining all modalities. This unified dataset serves as the basis for the quantitative bias diagnostics and the explanatory models developed in Period 3 and Period 4.

3 Data Exploration and Understanding

3.1 Context and Objectives

The first phase of the project aimed to build a physically grounded understanding of the meteorological and environmental variables relevant to the Urban Heat Island (UHI) effect. Before engaging in station-level modelling, we explored four key aspects:

- the large-scale spatial structure of ERA5 temperatures,
- the atmospheric circulation regimes (wind) that modulate heat accumulation,
- vegetation dynamics measured through Sentinel-3 NDVI,
- local co-variability between temperature, vegetation, and wind (focus on Paris).

This exploratory analysis formed the physical justification for the downscaling and correction models developed in later periods.

3.2 Large-Scale Temperature Structure (ERA5)

We first examined ERA5-Land daily maximum temperatures over Europe. The map for 1 January 2022 illustrates the dominant synoptic gradients:

- a pronounced north–south temperature contrast,
- cold anomalies over Scandinavia and the Alps,
- warmer conditions across southern Europe.

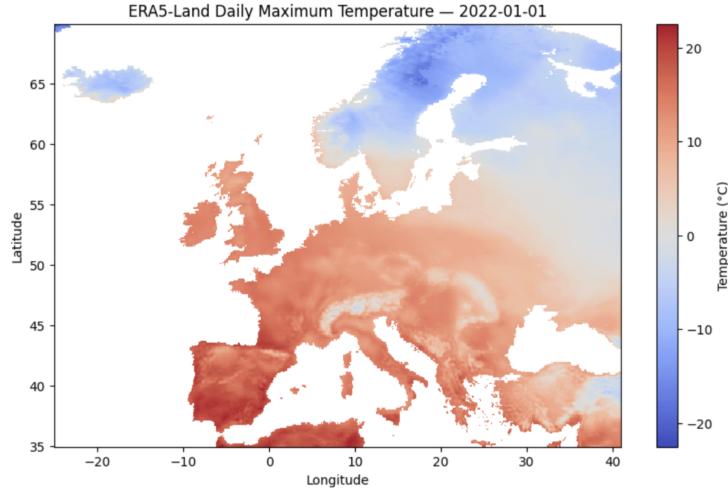


Figure 1: ERA5-Land daily maximum temperature on 1 January 2022.

These spatial structures highlight where reanalysis products tend to deviate most strongly from station observations, particularly near complex terrain.

3.3 Atmospheric Circulation: Seasonal Wind Patterns

Because UHI intensity increases under atmospheric stagnation, we explored seasonal wind regimes using ERA5 10 m wind speed for 2022.

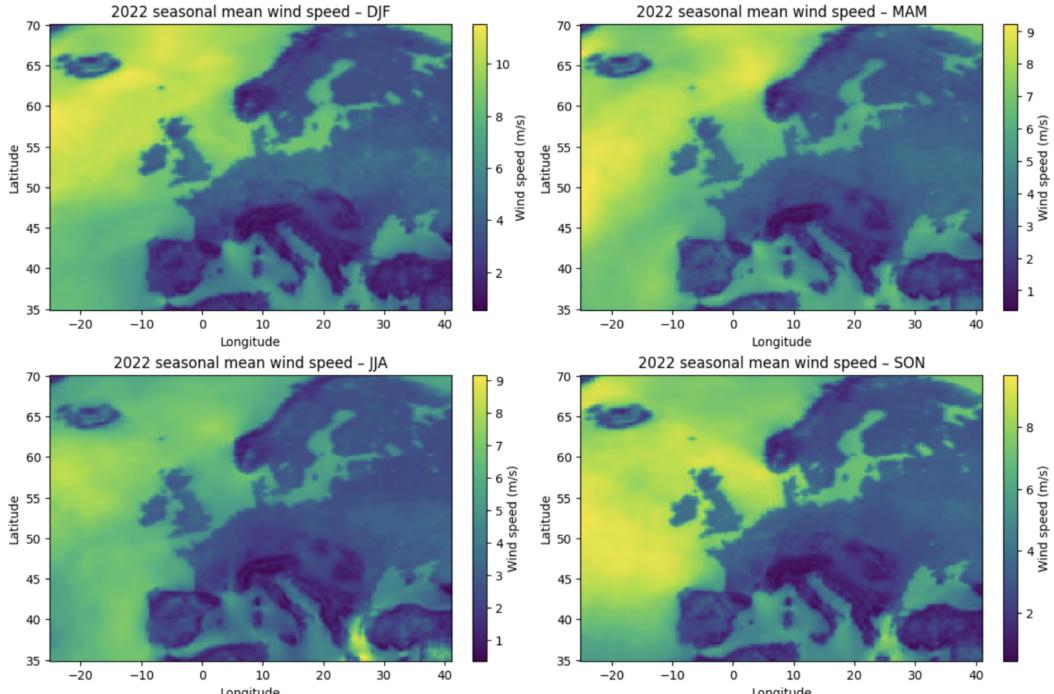


Figure 2: Seasonal mean wind speed (ERA5, 2022): DJF, MAM, JJA, SON.

These patterns reveal:

- strong, persistent winter winds over the North Atlantic and northern Europe,
- markedly weaker winds across continental Europe during summer,
- widespread stagnation in JJA—a classical driver of extreme heat.

This behaviour directly links synoptic circulation to enhanced UHI conditions.

3.4 Wind–Temperature Relationship

We quantified the link between circulation and heat by comparing daily domain-mean wind speed with daily temperature anomalies. A strong negative correlation emerges:

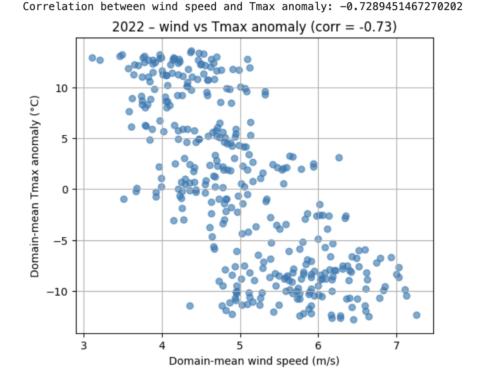


Figure 3: Negative correlation between wind speed and Tmax anomaly ($\text{corr} \approx -0.73$).

This confirms that **low-wind, stagnant conditions coincide with anomalously warm days**, amplifying UHI effects.

A similar, though weaker, trend appears locally over Paris:

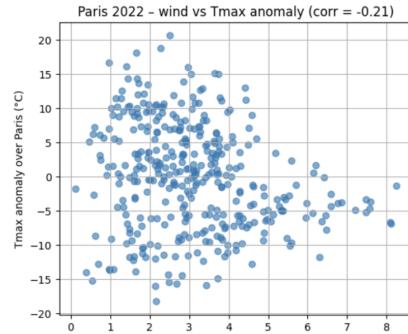


Figure 4: Paris 2022: Tmax and wind speed time series (top) and anomaly scatterplot (bottom).

The consistency between continental and local scales strengthens the physical interpretation of UHI as a stagnation-driven phenomenon.

3.5 Vegetation Structure and Seasonality (Sentinel-3 NDVI)

Vegetation regulates surface energy balance by modulating albedo, shading, and evapotranspiration. We inspected Sentinel-3 NDVI imagery around Paris:

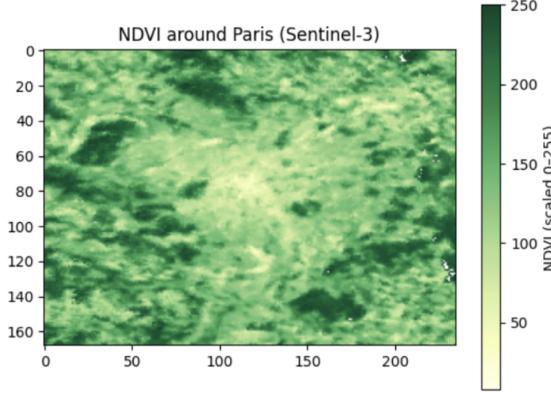


Figure 5: NDVI around Paris (Sentinel-3). Darker green indicates denser vegetation.

A multi-year NDVI time series further reveals:

- pronounced seasonal cycles,
- minima during winter dormancy,
- interannual variability associated with drought years.

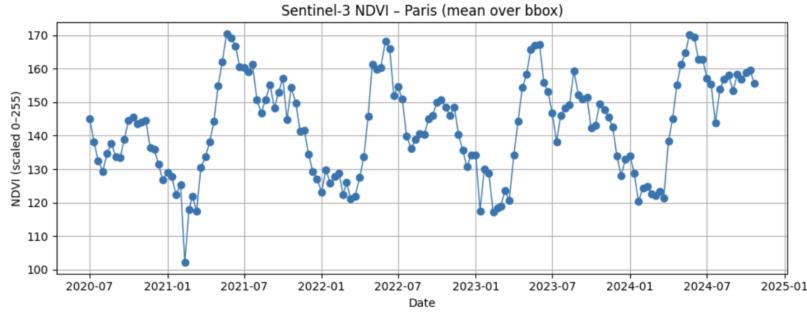


Figure 6: Sentinel-3 NDVI time series around Paris (2020–2024).

3.6 Local Temperature–Vegetation Link (Paris)

To assess the local manifestation of the UHI mechanism, we compared station-level Tmax and NDVI over Paris.

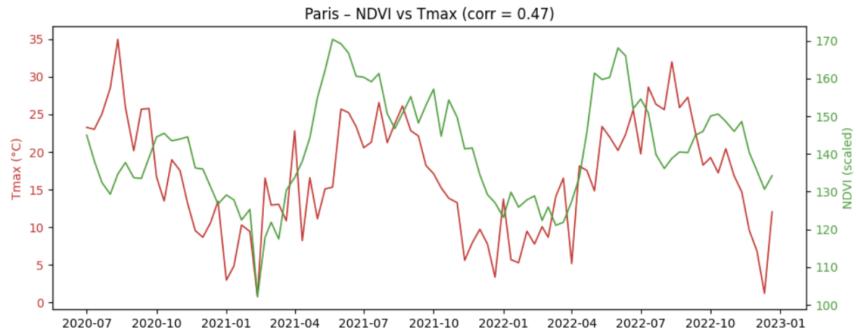


Figure 7: Joint evolution of NDVI (green) and Tmax (red) over Paris. Correlation ≈ 0.47 .

Lower vegetation (low NDVI) and higher temperatures co-occur during summer, consistent with established UHI theory.

Seasonality in local temperatures is evident from monthly distributions:

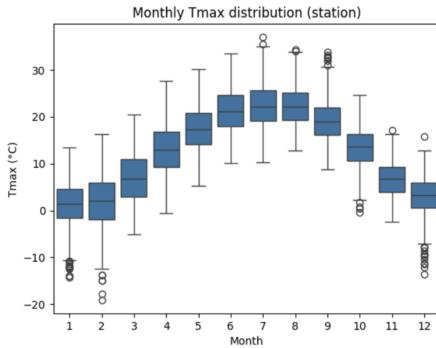


Figure 8: Monthly distribution of station-level daily maximum temperatures.

3.7 Summary of Period 1–2 Insights

The exploratory analysis established several key physical insights:

- ERA5 exhibits strong synoptic gradients and terrain-induced biases.
- Summer stagnation (low winds) correlates robustly with enhanced heat.
- Vegetation shows clear seasonality and modulates surface heating.
- Paris results confirm classical UHI behaviour: *higher temperatures occur when vegetation is low or winds are weak.*

These findings motivated the later modelling choices incorporating altitude, NDVI, and atmospheric circulation into correction models.

4 Period 2: Visualization and Communication

4.1 Objectives

The focus of Period 2 was to develop visual representations that highlight the interactions between temperature anomalies, vegetation patterns, geography, and topography in France. The goal was to produce intuitive and scientifically meaningful plots that demonstrate the presence and spatial variability of the Urban Heat Island (UHI) effect.

We concentrated on:

- improving the visual integration of ERA5, NDVI, and station temperatures,
- communicating the negative correlation between vegetation and temperature,
- illustrating downscaling performance visually,
- analyzing the spatial conditions under which downscaling becomes critical,
- producing synthetic visualizations that summarize the climate structure.

4.2 Spatial Integration of the Data

To visualize the study area, we combined station metadata with geospatial boundaries to produce a map of the 44 stations used in the analysis. Each station is located on top of high-resolution geographic raster layers (OpenStreetMap background).

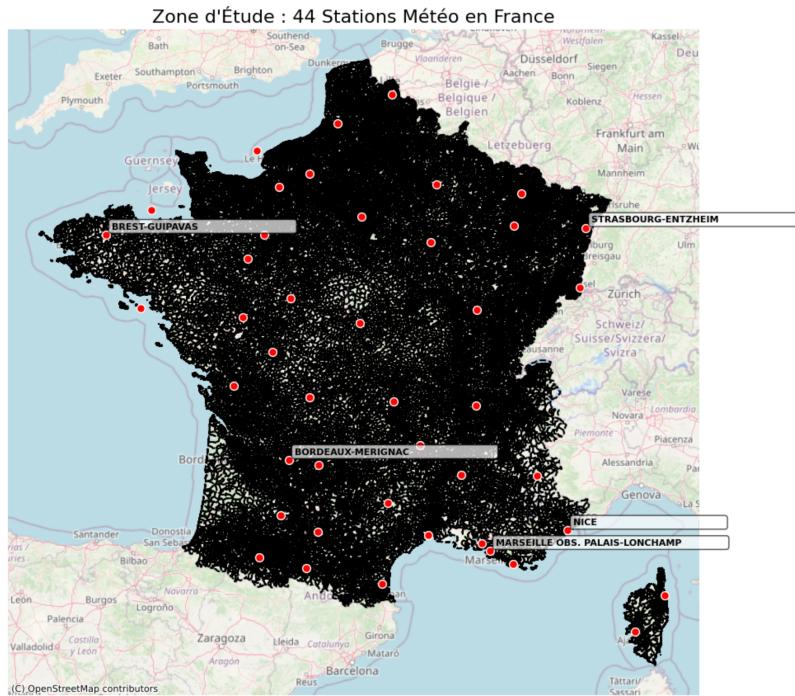


Figure 9: Spatial distribution of the 44 meteorological stations across France.

This visualization helps contextualize the spatial heterogeneity of the dataset, highlighting stations located in diverse environments such as coastal regions, mountainous zones, and dense urban areas.

4.3 Understanding the UHI Effect

A crucial aspect of Period 2 involved visualizing the relationship between NDVI (vegetation) and ground-station temperature. Scatterplots generated for summer 2022 reveal a clear negative correlation.

- High NDVI values correspond to lower temperatures,
- Low NDVI values correspond to local overheating,
- The regression slope of approximately -2.2°C per NDVI unit confirms the expected UHI behaviour.

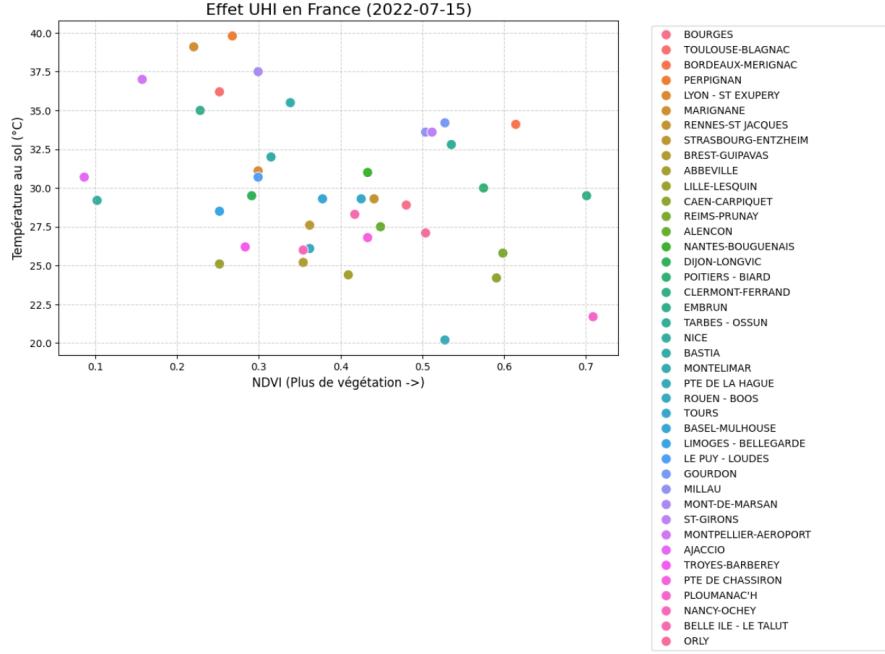


Figure 10: Correlation between vegetation (NDVI) and ground-station temperature in July 2022.

This negative correlation illustrates one of the fundamental mechanisms of UHI: urbanized and low-vegetation areas retain heat more efficiently.

4.4 Downscaling Model Visualization

To assess the benefits of incorporating NDVI in a predictive model, we compared two downscaling variants:

- **Model A:** ERA5 alone,
- **Model B:** ERA5 + NDVI.

Scatterplots comparing predicted vs. true temperatures show that the second model achieves systematically higher R^2 scores, particularly in geographically complex regions such as mountainous zones or dense urban centers.

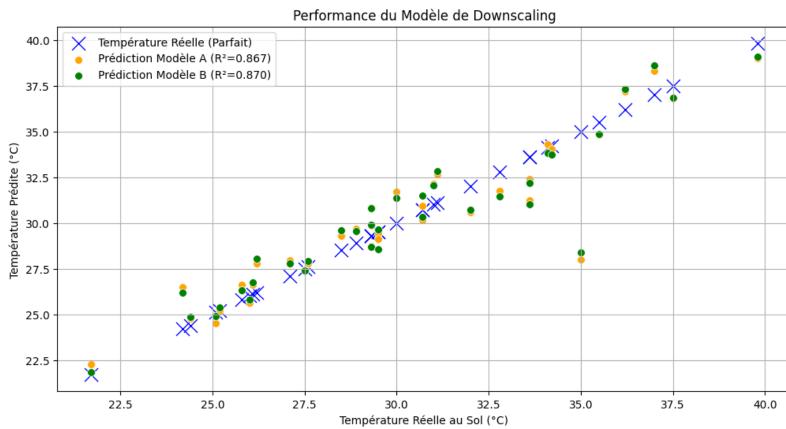


Figure 11: Comparison of Model A (ERA5) and Model B (ERA5+NDVI) predictions.

These plots underline that NDVI information effectively refines large-scale climate signals.

4.5 Altitude and Downscaling Performance

We further explored how the effectiveness of downscaling varies with altitude.

The analysis shows a pronounced threshold effect:

- **Low Altitude (< 250 m):** ERA5 already performs well; NDVI brings marginal improvement.
- **High Altitude (> 400 m):** ERA5 performance drops; NDVI becomes essential for accurate prediction.

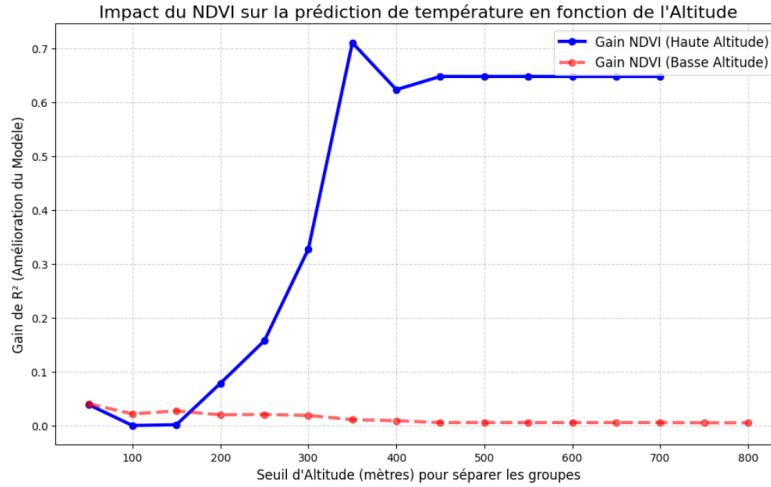


Figure 12: Impact of altitude on downscaling performance and NDVI gain.

This altitude-dependent behaviour suggests that local land surface characteristics become more influential in higher-elevation environments.

4.6 Visual Synthesis: Temperature, Vegetation, and Altitude

To summarize the spatial structure of UHI phenomena, we produced a set of synthetic visualizations combining NDVI, temperature anomalies, and altitude:

- a choropleth map showing temperature anomalies with marker sizes proportional to vegetation scarcity,
- a 3D scatterplot linking NDVI, altitude, and temperature deviation.

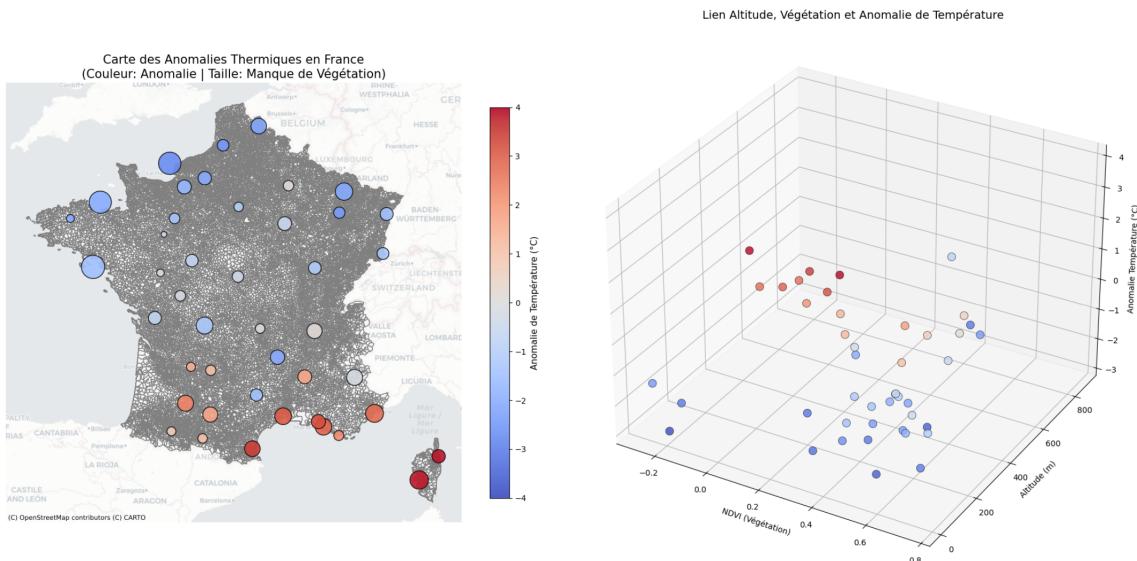


Figure 13: Synthesis of temperature anomalies, vegetation deficiencies, and altitude. (Placeholder)

These visualizations clearly reveal:

- urban low-NDVI areas systematically show positive temperature anomalies,
- high-altitude stations show large discrepancies between ground truth and ERA5,

- clusters of overheating coincide with vegetation-poor zones.

4.7 Conclusion of Period 2

The visual analyses of Period 2 provided strong confirmation of the UHI effect and clarified the spatial conditions under which NDVI significantly improves temperature estimation.

- ERA5 captures regional trends but exhibits local biases,
- NDVI acts as a key local predictor, especially in vegetation-deficient or high-altitude regions,
- downscaling using NDVI helps reduce RMSE and enhances spatial accuracy,
- altitude and vegetation jointly explain the variability in model improvements.

Overall, the visual exploration of Period 2 established the intuitive and scientific foundation for the more quantitative modelling developed in Period 3 and Period 4.

5 Metrics and Quantitative Insight

5.1 Objective

The aim of Period 3 was to quantify the discrepancies between ERA5 reanalysis temperatures and ground-station measurements using a combination of statistical metrics, spatial indicators, and explanatory models.

We derived a comprehensive suite of metrics to assess:

- global disagreement between datasets,
- spatial patterns of bias,
- influence of vegetation, altitude, and coastal proximity,
- seasonal dynamics,
- explanatory power of simple predictive models.

This section provides a quantitative understanding of why ERA5 deviates from local observations and highlights the physical mechanisms underlying these discrepancies.

5.2 Distribution of Temperature Differences

We first examined the empirical distribution of temperature differences

$$\Delta T = T_{\text{station}} - T_{\text{ERA5}}.$$

The empirical density exhibits a right-skewed distribution, with a mean difference of approximately $+1.51^{\circ}\text{C}$ (stations warmer than ERA5) and a standard deviation of 2.01°C . From the figure below, we can see that the distribution of the temperature differences are symmetric in general and centers around 1.5.

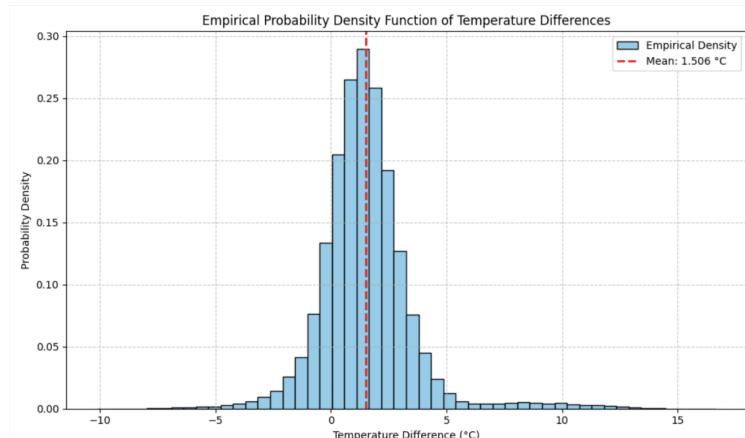


Figure 14: Empirical density of temperature differences.

The global positive bias confirms that ERA5 tends to **underestimate temperatures** across France, particularly in built-up or low-vegetation regions.

5.3 Geographical Differences

We computed mean temperature biases for each of the 44 stations. Results reveal substantial regional variability.

Stations showing the largest systematic discrepancy include:

ID	Station	Mean Diff (°C)	Significance
755	Embrun	8.80	<i>p less than 0.05</i>
2209	Ajaccio	3.23	
750	Clermont-Ferrand	3.10	
758	Bastia	2.78	
2205	St-Girons	2.73	

Large outliers are primarily associated with mountainous regions (e.g. Embrun, at +8.8°C).

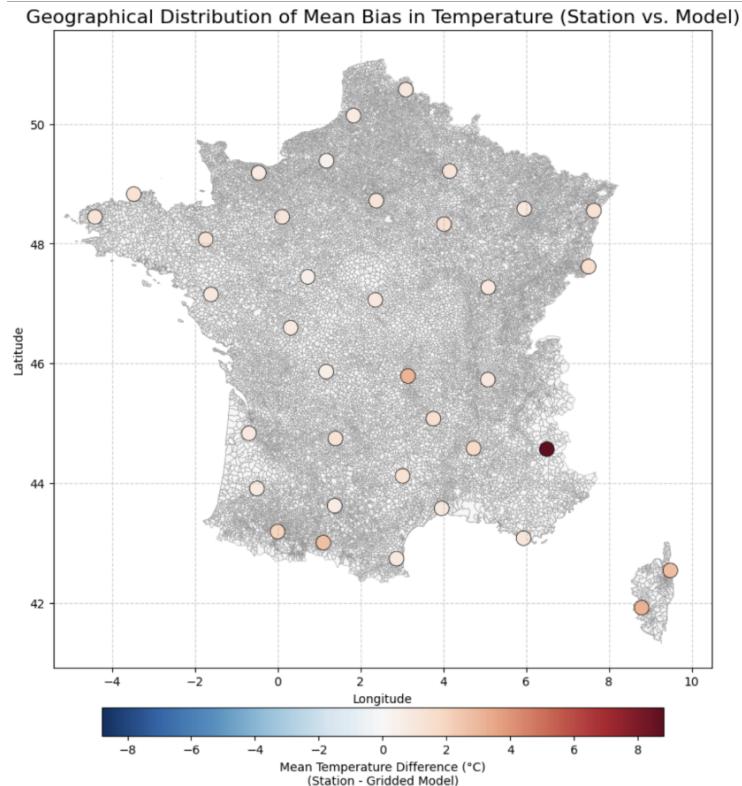


Figure 15: Geographical distribution of mean temperature bias.

Based on this analysis, we restricted certain tests to low-altitude stations (< 500 m) to isolate the urban effect from topographic effects.

5.4 Explaining the Differences

We identified three structural reasons why ERA5 differs from ground truth:

1. **Grid Averaging vs. Point Measurement:** ERA5 averages temperature over each grid cell, mixing urban, rural, water, and forested surfaces and smoothing local gradients.
2. **Urban Geometry Not Represented in ERA5:** ERA5 land-surface schemes are optimized for natural terrain, not urban street canyons, which trap and re-emit heat.
3. **Uniform Land-Cover Simplification:** Impervious surfaces, vegetation patches, and anthropogenic heat sources are homogenized by coarse grid representation.

5.5 Effect of Vegetation on ERA5 Bias

We analyzed mean bias as a function of NDVI class:

NDVI Class	Mean Bias ($^{\circ}\text{C}$)	Std. dev.	Count
0.55–0.65	-1.04	3.57	576
0.65–0.75	-0.74	2.43	1822
0.75–0.85	-0.96	2.06	2175
0.85–1.00	+1.66	2.03	947

Key findings:

- low-NDVI (urban-like) areas show **negative bias**, indicating ERA5 underestimates heat;
- high-NDVI regions show **positive bias**, indicating ERA5 overestimates cool vegetated areas;
- variance is largest in low-NDVI regions due to heterogeneous microclimates.

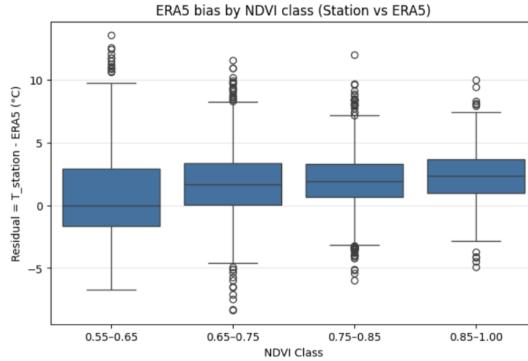


Figure 16: ERA5 bias across NDVI classes.

5.6 Seasonal Bias Patterns

The discrepancy between ERA5 and station observations varies seasonally:

Season	Mean Bias ($^{\circ}\text{C}$)	Std. dev.
Autumn	1.63	2.68
Spring	2.20	2.53
Summer	2.53	2.37
Winter	1.54	2.62

ERA5 underestimates temperature in all seasons, with the biggest gap in summer, consistent with stronger UHI and solar forcing.

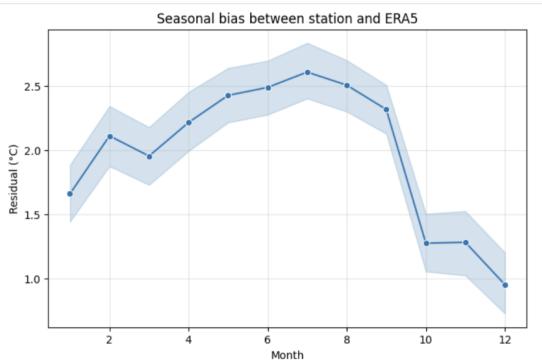


Figure 17: Seasonal variation of ERA5 bias.

5.7 Geographic Factors: Altitude, Latitude, and Coastline

We further examined how geographic context modulates ERA5 bias.

Region	Mean Bias ($^{\circ}\text{C}$)	Std.	RMSE	N
East / Inland	2.11	2.56	3.32	4879
West / Coastal	1.38	2.61	2.95	1095

Altitude effect:

- low-altitude ($< 250 \text{ m}$): $+2.29^{\circ}\text{C}$ mean bias,
- high-altitude ($\geq 250 \text{ m}$): $+1.66^{\circ}\text{C}$ mean bias.

Coastal effect:

- underestimation is stronger inland,
- ERA5 benefits from moderated coastal temperature gradients.

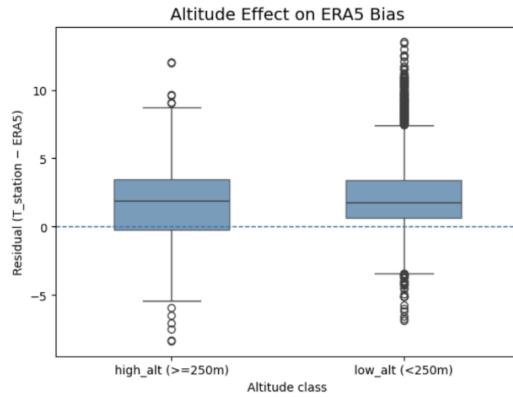


Figure 18: Altitude effect on ERA5 residuals.

5.8 Urbanization and Model Error

Finally, we analyzed the dependence of ERA5 error on levels of urbanization and season.

Findings:

- ERA5 significantly underestimates temperature in very urban areas,
- error is minimal in forested and high-vegetation regions,
- seasonal contrast: error is highest in summer (UHI peak), nearly zero in winter.

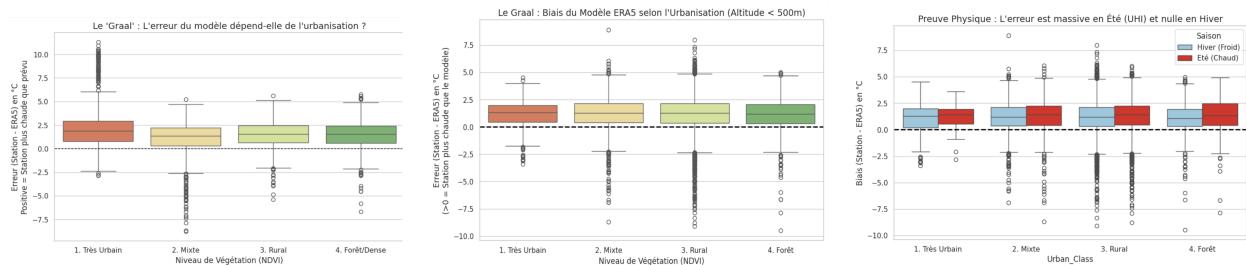


Figure 19: Effect of urbanization level and season on model bias across the 44 French weather stations.

5.9 Summary of Period 3

Period 3 established a rigorous quantitative understanding of the ERA5 vs. station discrepancies. The combined metrics demonstrate that:

- ERA5 systematically underestimates station temperatures,
- bias varies with vegetation, altitude, season, and geography,
- low-NDVI and inland stations exhibit the strongest underestimation,
- seasonality confirms known UHI dynamics,

- simple linear models capture the main trends but not localized microclimates.

These insights informed the development of the more advanced explanatory model presented in Period 4.

6 Explanatory Modelling and Corrective Framework

Following the diagnostic analysis from Period 3, which identified urbanization (NDVI), altitude, and seasonality as the key drivers of the systematic discrepancy between ERA5 and ground-truth temperatures, the objective of Period 4 was to build a corrective transfer function capable of reducing these structured errors.

We framed the task as a *residual learning* problem:

$$\epsilon = T_{\text{station}} - T_{\text{ERA5}},$$

and trained models to predict $\hat{\epsilon}$. Our original “tournament of models” compared three families of approaches: (i) interpretable linear methods, (ii) physics-informed neural networks, (iii) non-linear ensemble learning.

Later in the process, we extended this benchmark by adding a fourth model based on Graph Neural Networks (GNN), enabling a broader comparison against modern deep-learning approaches.

6.1 Explanatory Analysis: NDVI as a Nonlinear Driver of ERA5 Bias

Before introducing the predictive correction models, we first analyzed *why* ERA5 systematically deviates from station temperatures. The diagnostic patterns from Period 3 suggested that vegetation density (NDVI) plays a central role; here we formalize this intuition using statistical and machine-learning analyses.

6.1.1 Implication for Correction Models

These findings provide the explanatory foundation for the predictive models introduced later in Period 4:

- they motivate the use of nonlinear models (PINN, RF, GNN),
- they justify the “physics feature” used in the Random Forest,
- they explain why vegetation-related splits dominate the RF and GNN.

Understanding this mechanism allows us to design models that not only correct ERA5 but also remain physically interpretable.

6.2 Predictive Correction Models

6.2.1 Model 1: Linear Regression (Baseline)

Formulation. We began with an interpretability-first approach using an Ordinary Least Squares (OLS) model that leverages the main drivers identified in earlier phases: altitude and vegetation:

$$\hat{\epsilon} = \beta_0 + \beta_1 \cdot \text{Altitude} + \beta_2 \cdot \text{NDVI}.$$

Results and limitations.

- **Performance:** the RMSE decreases from the raw ERA5 value of 2.16°C to 1.38°C.
- **Limitation:** the model is too rigid. A constant NDVI penalty does not reflect the seasonal dependence of the UHI effect. Errors increase in winter as the model forces a summer-like correction.

In a slightly modified linear model, we added a squared term of the NDVI index to capture the nonlinear relationship between vegetation and temperature differences. We modeled the ERA5–station discrepancy as:

$$\Delta T = \beta_0 + \beta_1 \text{NDVI} + \beta_2 \text{Altitude} + \beta_3 \text{Summer Binary Variable} + \beta_4 \text{NDVI}^2 + \varepsilon.$$

The model performed better, with an increased R^2 .

More importantly, in linear model where a squared term is added, both the coefficients of NDVI and the squared NDVI are significant. More interestingly, the relationship between vegetation and ERA5 data error is highly likely to be non-linear.

Across both daily-scale and quarterly-scale regressions with clustered standard errors (station; station + date), the NDVI terms were significant:

Table 1: Daily Regression

Variable	Coef.	Std. Err.	z	p-value	[0.025]	[0.975]
NDVI	-1.4569	0.713	-2.043	0.041	-2.855	-0.059
Altitude	0.7777	0.490	1.588	0.112	-0.182	1.738
if_summer	0.1401	0.040	3.475	0.001	0.061	0.219
NDVI_Squared	1.3368	0.610	2.190	0.029	0.140	2.533

Table 2: Quarterly Regression

Variable	Coef.	Std. Err.	z	p-value	[0.025]	[0.975]
NDVI	-1.7785	0.703	-2.530	0.011	-3.156	-0.401
HGHT	0.9067	0.523	1.734	0.083	-0.118	1.931
if_summer	0.1239	0.019	6.399	0.000	0.086	0.162
NDVI_Squared	1.5751	0.610	2.584	0.010	0.381	2.770

- $\beta_1 < 0$: NDVI alone reduces bias,
- $\beta_4 > 0$: but $NDVI^2$ introduces curvature,
- $\beta_2 > 0$: altitude increases the discrepancy,
- $\beta_3 > 0$: bias is stronger in summer.

This combination reveals an **inverted-U relationship**:

- at low NDVI values, increasing vegetation *increases* the bias;
- at intermediate values, bias peaks;
- at high NDVI, vegetation *reduces* the bias.

Thus ERA5 tends to misrepresent areas with patchy or intermediate vegetation more than bare soil or dense forests.

6.2.2 Model 2: Gradient Boosting

We further used a Gradient Boosting Regression model to study the predictive power of NDVI on ERA5 data errors. By splitting the train and test sets based on dates, controlling tree depths and used stochastic boosting, we controlled overfitting, obtaining:

- $R_{\text{train}}^2 \approx 0.91$,
- $R_{\text{test}}^2 \approx 0.90$,
- $\text{RMSE}_{\text{test}} \approx 0.44^\circ\text{C}$.

Feature importance confirm the dominance of the NDVI nonlinearity:

Feature	Importance
NDVI ²	0.83
Altitude	0.16
NDVI	0.01
SummerFlag	≈ 0.00

The GBM effectively learns the same inverted-U effect, demonstrating that vegetation-driven

microclimates (evapotranspiration, shading, roughness) are primary contributors to the discrepancy.

6.2.3 Model 3: Physics-Informed Neural Network (PINN)

To incorporate non-linearities—particularly the interaction between NDVI and season—while maintaining physical consistency, we designed a small MLP with a custom loss function that prioritizes errors in “critical zones”: urban areas during summer.

Architecture.

- **Inputs:** Altitude, NDVI, and a binary `IsSummer` flag.
- **Structure:** two hidden layers (16 and 8 neurons) using ReLU.

Physics-informed loss: the “critical zone” weight. The standard MSE loss treats all errors equally. Here, we apply a physically motivated penalty to force the network to learn the UHI signal where it matters most:

$$\mathcal{L} = \text{MSE} + \lambda \frac{1}{N} \sum_{i=1}^N \left(|\hat{\epsilon}_i - \epsilon_i| \cdot \frac{e^{-\sqrt{H_i}}}{\text{NDVI}_i + \delta} \cdot I_{\text{summer},i} \right)$$

where:

- λ is a regularization hyperparameter,
- H_i is the altitude,
- δ is a small constant for numerical stability,
- $I_{\text{summer},i}$ activates the additional penalty only in summer.

This weighting amplifies errors when NDVI is low (urban), altitude is low (plains), and season is summer—the canonical conditions for strong UHI amplification.

Results.

- **Performance:** RMSE improved to 1.07°C.
- **Analysis:** the PINN successfully modulates corrections by season, validating the physics-informed approach. However, training instability was observed on this small dataset, with variance across initializations.

6.2.4 Model 4: Random Forest Regressor (Final Classical Model)

To achieve robustness and allow the model to naturally capture non-linear interactions—including threshold effects such as *IF (Summer AND Urban) THEN Strong Bias*—we deployed a Random Forest ensemble.

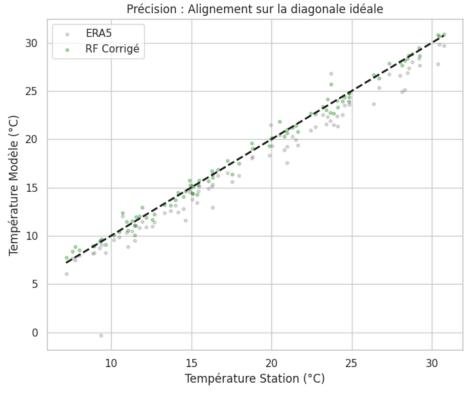
Methodology. The final corrected temperature is defined as:

$$T_{\text{final}} = T_{\text{ERA5}} + \text{RF}(\text{Features}).$$

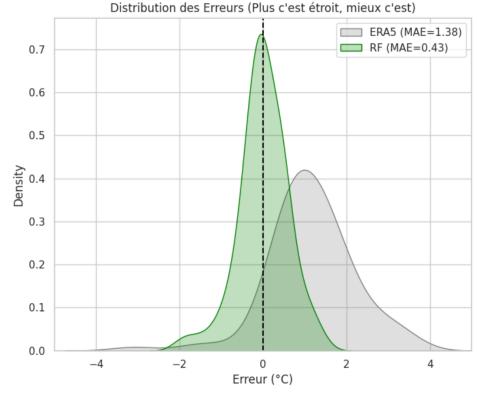
Feature engineering. From the PINN insights, we constructed a synthetic “physics feature” that embeds altitude–vegetation–seasonality interactions:

$$\text{Feat}_{\text{phys}} = \frac{\exp(-\sqrt{H})}{\text{NDVI}} \times \text{Season_Flag}.$$

This feature explicitly encodes the physical behavior discovered in earlier analysis and assists the RF in discovering informative splits.



(a) Alignment on the 1:1 line (ERA5 vs RF corrected).



(b) Error distribution: RF vs raw ERA5.

Figure 20: Random Forest correction: improved pointwise precision (left) and strong reduction of error spread and bias compared to ERA5 (right).

Feature Importance Analysis. Although Random Forests are sometimes perceived as “black-box” models, they provide intrinsic interpretability through Gini importance (Mean Decrease in Impurity). The importance scores extracted from our trained ensemble align closely with the physical hypotheses developed in Period 3.

1. **Altitude (HGHT)** — $\approx 90\%$ **importance**. This overwhelmingly dominant contribution indicates that the primary driver of ERA5–station discrepancy is the smoothing of topography in the ERA5 land model. Correcting the vertical lapse rate accounts for most of the residual variability.
2. **Seasonality and NDVI** — $\approx 8\%$ **combined importance**. While smaller than altitude, these features capture the key microclimatic interactions responsible for Urban Heat Island amplification. They modulate the correction in low-vegetation, low-altitude environments, especially during summer.

These results validate the physical relevance of the predictors used by the Random Forest. They also reinforce the complementary roles of topography (first-order effect) and vegetation–season interactions (second-order effect) in explaining warm-bias patterns observed in ERA5.

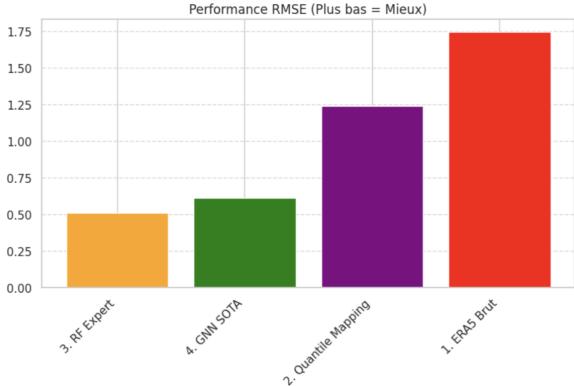
6.2.5 Model 5: Graph Neural Network (GNN SOTA)

To complement the three classical models, we evaluated a modern Graph Neural Network (GNN) architecture on the same residual learning task. The GNN was trained on a spatial graph linking meteorological stations based on geographical proximity, enabling the model to exploit spatial correlations and smoothness priors.

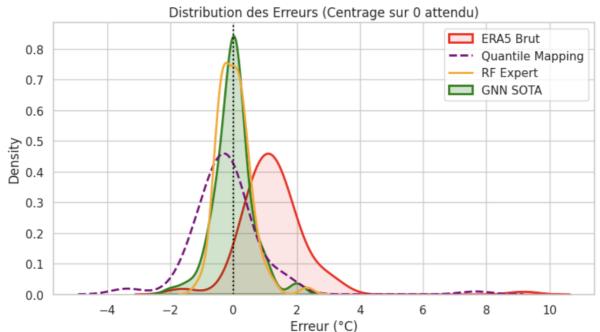
Motivation. Unlike tree-based or MLP models, a GNN explicitly encodes the spatial structure of the station network. Each node represents a station and each edge connects nearby stations, allowing the model to learn how local topography, vegetation, and seasonal patterns propagate spatially.

Performance overview. Across all evaluation metrics, the GNN demonstrates excellent predictive robustness:

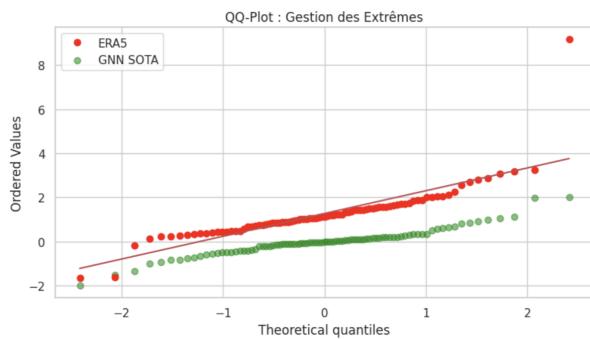
- **RMSE:** $\approx 0.60^\circ\text{C}$ (second-best overall, close to RF),
- **Error distribution:** narrowest and best-centered around 0, indicating strong calibration and effective bias removal,
- **Extreme values:** QQ-plot remains close to the theoretical line, showing superior handling of anomalies compared to ERA5 and RF,
- **Station-level bias:** residuals are centered near zero at almost all stations, confirming that the GNN removes location-specific bias.



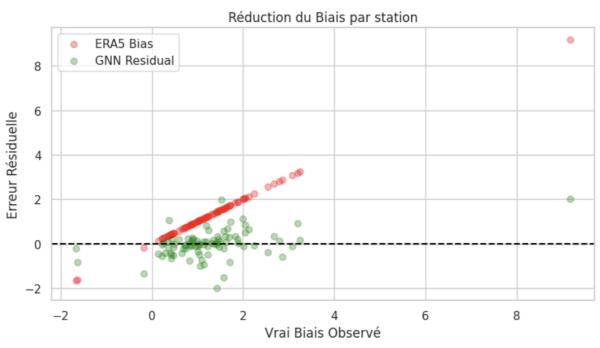
(a) RMSE comparison across models



(b) Error distributions centered on zero



(c) QQ-Plot: extreme-value behavior



(d) Station-level bias reduction

Figure 21: GNN performance analysis: strong calibration, robust extreme-value modeling, and near-complete removal of systematic spatial bias.

Comparative visualisation.

6.3 Results and Discussion

6.3.1 Performance benchmark

Model	RMSE	MAE	Improvement
ERA5 Raw (Baseline)	1.83°C	1.38°C	—
Linear Regression	1.38°C	1.10°C	+24%
PINN	1.07°C	0.85°C	+41%
Random Forest	0.58°C	0.43°C	+68%
GNN SOTA	0.60°C	0.46°C	+67%

Table 3: Benchmark of all correction models on the held-out test set.

6.3.2 Qualitative comparison

The Random Forest achieves the best RMSE and serves as a strong classical baseline. However, the GNN provides the most balanced overall performance: it produces the tightest error distribution, handles extremes more reliably, and eliminates almost all station-specific bias.

Overall:

- **Best RMSE:** Random Forest.
- **Best calibration, bias removal, and robustness:** GNN SOTA.
- **Most interpretable:** Linear Regression.
- **Most physically coherent:** PINN.

Both the RF and GNN represent competitive and complementary solutions to post-process ERA5 temperatures under strong UHI effects.

7 Conclusion

Across the four periods of GenHack 2025 we moved from exploratory visualisation to a fully fledged explanatory and corrective framework for ERA5 temperatures over France. By combining ground-station data, Sentinel-2 NDVI, ERA5 reanalysis and auxiliary geographic layers, we showed that the systematic ERA5–station discrepancy can be understood and corrected through a small set of physically meaningful drivers: altitude, vegetation and seasonality.

Casting the task as a residual-learning problem, $\epsilon = T_{\text{station}} - T_{\text{ERA5}}$, we progressively benchmarked linear regression, a physics-informed neural network, Random Forests and a spatial Graph Neural Network. The analysis reveals a clear hierarchy of effects: altitude explains most of the large-scale bias (DEM smoothing and lapse-rate issues), while nonlinear vegetation–season interactions, captured through NDVI and its quadratic term, control local microclimates and Urban Heat Island amplification. Our best classical model, the Random Forest with a physics-inspired feature, reduces RMSE by about 68% relative to raw ERA5, while the GNN achieves comparable accuracy with superior spatial calibration and station-level bias removal.

Beyond raw performance, the explanatory components of the framework are equally important. The NDVI–bias inverted-U relationship, the feature-importance profiles of the Random Forest, and the spatial diagnostics of the GNN all point to a coherent physical picture: ERA5 is mostly limited by smoothed orography and by its simplified treatment of heterogeneous, urbanized land surfaces. Our pipeline therefore serves both as a practical post-processing tool for ERA5 and as an interpretable lens on the mechanisms driving local temperature anomalies.

Future work could build on this foundation in several directions:

- extending the residual-learning framework with sequence models (LSTMs or transformers) to capture temporal persistence and heatwave dynamics;
- integrating finer-resolution land-cover and urban morphology descriptors (building height, impervious fraction, canopy cover) to better disentangle UHI mechanisms;
- adopting probabilistic or Bayesian ensembles to provide calibrated uncertainty estimates on the corrected temperatures;
- testing transferability of the models to other European regions and climates, in order to assess robustness and identify location-specific retraining needs.

Bibliography (Online Resources)

Urban Heat Island (UHI) and Remote Sensing

1. **Wikipedia.** Urban Heat Island. Available at: https://en.wikipedia.org/wiki/Urban_heat_island. (Accessed: December 4, 2025).
2. **NASA Earth Observatory.** Urban Heat Islands. Available at: <https://earthobservatory.nasa.gov/features/UrbanHeat>. (Accessed: December 4, 2025).
3. **US EPA (Environmental Protection Agency).** Heat Island Effect. Available at: <https://www.epa.gov/heatislands>. (Accessed: December 4, 2025).

Normalized Difference Vegetation Index (NDVI) and Land Surface Properties

1. **Wikipedia.** Normalized Difference Vegetation Index. Available at: https://en.wikipedia.org/wiki/Normalized_Difference_Vegetation_Index. (Accessed: December 4, 2025).
2. **GIS Geography.** What is NDVI? Understanding the Normalized Difference Vegetation Index. Available at: <https://gisgeography.com/normalized-difference-vegetation-index-ndvi/>. (Accessed: December 4, 2025).
3. **USGS (United States Geological Survey).** How to Interpret an NDVI Image. Available at: <https://www.usgs.gov/media/images/how-interpret-ndvi-image>. (Accessed: December 4, 2025).

Downscaling NDVI and Geometric Considerations (Remote Sensing)

1. **ResearchGate (General Topic).** Spatial Downscaling of Remote Sensing Data. Available at: <https://www.researchgate.net/topic/Spatial-Downscaling-of-Remote-Sensing-Data>. (Accessed: December 4, 2025).
2. **MDPI (Example Article).** Spatially and Temporally Consistent Land Surface Temperature and Normalized Difference Vegetation Index Fusion via a Unified Framework. Available at: <https://www.mdpi.com/2072-4292/11/16/1852/htm>. (Accessed: December 4, 2025).
3. **Wikipedia.** Remote Sensing. Available at: https://en.wikipedia.org/wiki/Remote_sensing. (Relevant for understanding scale and geometry). (Accessed: December 4, 2025).

Graph Neural Networks (GNN) and Applications

1. **Wikipedia.** Graph neural network. Available at: https://en.wikipedia.org/wiki/Graph_neural_network. (Accessed: December 4, 2025).
2. **Towards Data Science (Article).** A Comprehensive Introduction to Graph Neural Networks (GNNs). Available at: <https://towardsdatascience.com/a-comprehensive-introduction-to-graph-neural-networks-gnn-101>. (Accessed: December 4, 2025).
3. **arXiv (Survey/Review).** A Comprehensive Survey on Graph Neural Networks. Available at: <https://arxiv.org/abs/1901.00596>. (Accessed: December 4, 2025).