Team 23 NoName

Ayoub Olulad Ali
Maria Cristina Centorame
Yuxin Li

Discrepancies between Datasets

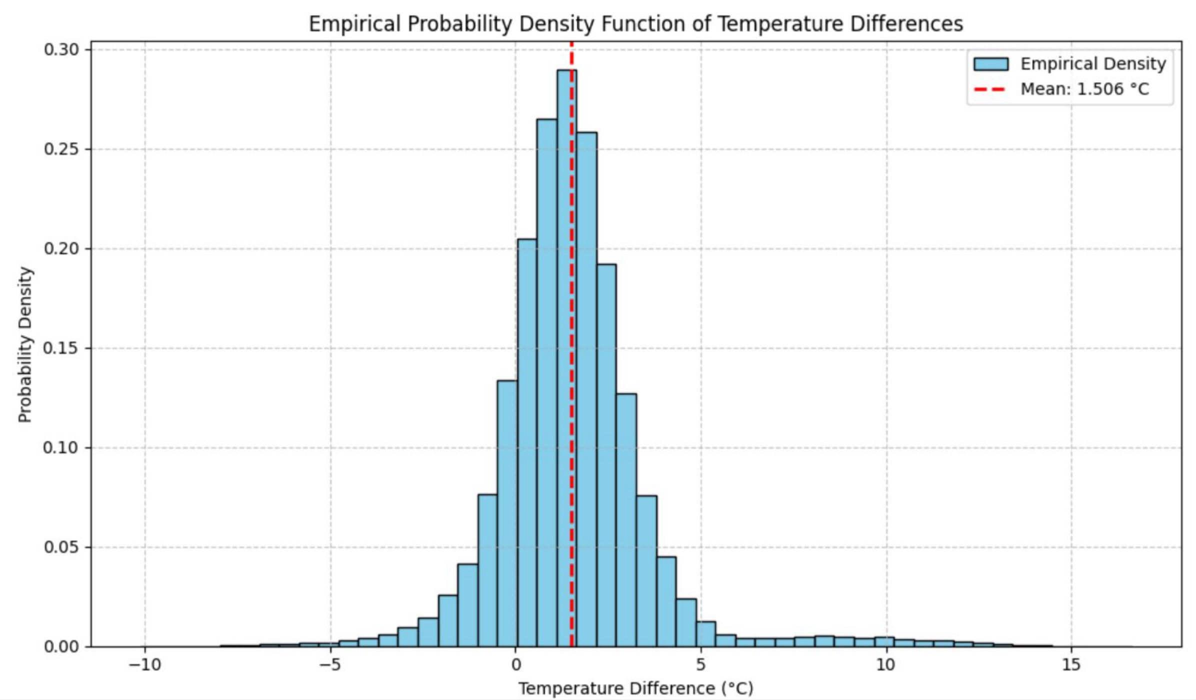
Comparing ERA5 and Weather Station Data

ERA5 vs Ground Truth

After matching the weather station temperature data with the nearest available ERA5 temperature data available in France, we found a difference:

Weather station data reports a higher temperature.

The mean of difference (weather station data - ERA5 data) ≈ 1.51, standard deviation is 2.01.



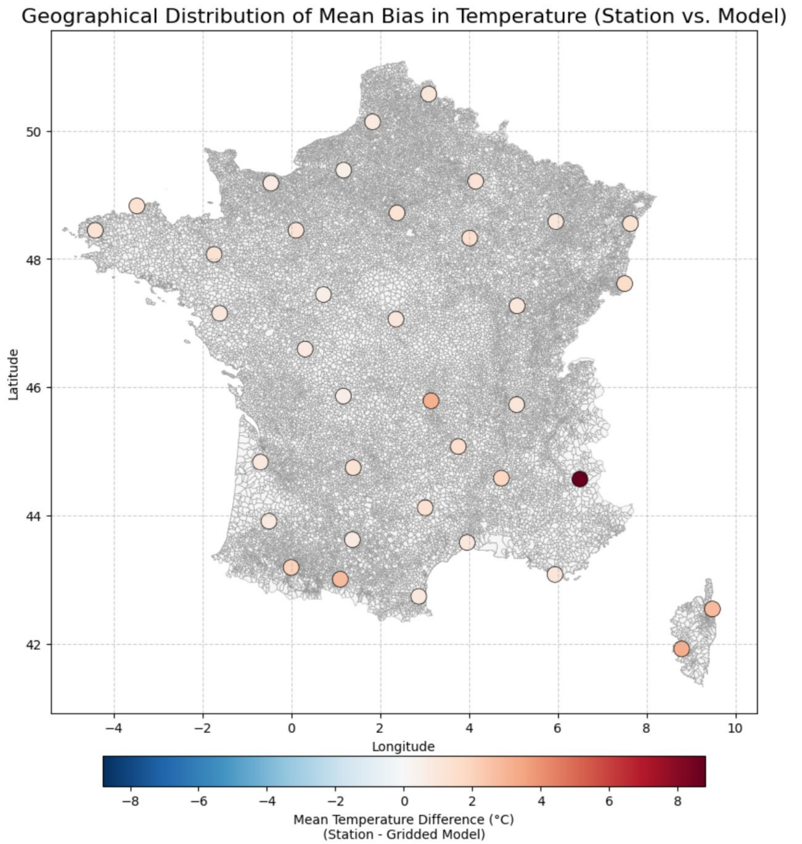
Geographical Differences:

By looking at temperature difference in all 44 weather stations in France, we found out the differences to be significant across all regions.

ID	Weather Station Name	Mean Temperature Diff	Significance
755	EMBRUN	8.80	Significant ($p < 0.05$)
2209	AJACCIO	3.23	
750	CLERMONT-FERRAND	3.10	
758	BASTIA	2.78	
2205	ST-GIRONS	2.73	

Top 5 Most Different Weather Stations

As shown on the right graph, the differences are more severe in the south-eastern region. Extreme outliers (e.g., Embrun +8.8° C) are driven by altitude/relief.
STRATEGIC DECISION: To capture the pure Urban Heat Island effect, we filtered the dataset to low-altitude stations (< 500m).

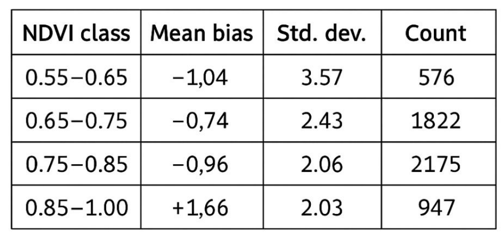

Geographical Distribution of Mean Bias in Temperature (Station vs. Model)
Latitude
50
48
46
44
42
Longitude
−4
−2
0
2
4
6
8
10
−8
−6
−4
−2
0
2
4
6
8
Mean Temperature Difference (°C)
(Station - Gridded Model)

NDVI & Urbanization Effect on ERA5 Bias

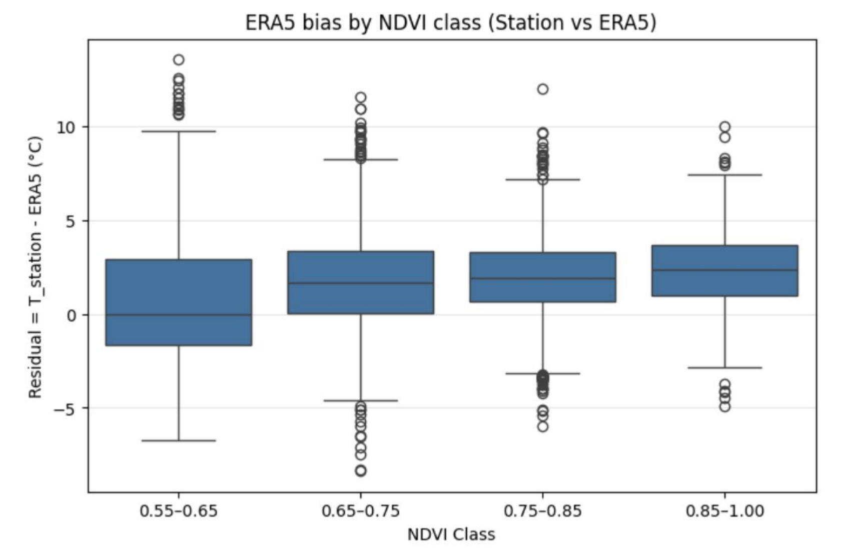
Vegetation Strongly Modulates ERA5 Temperature Errors

Key Results

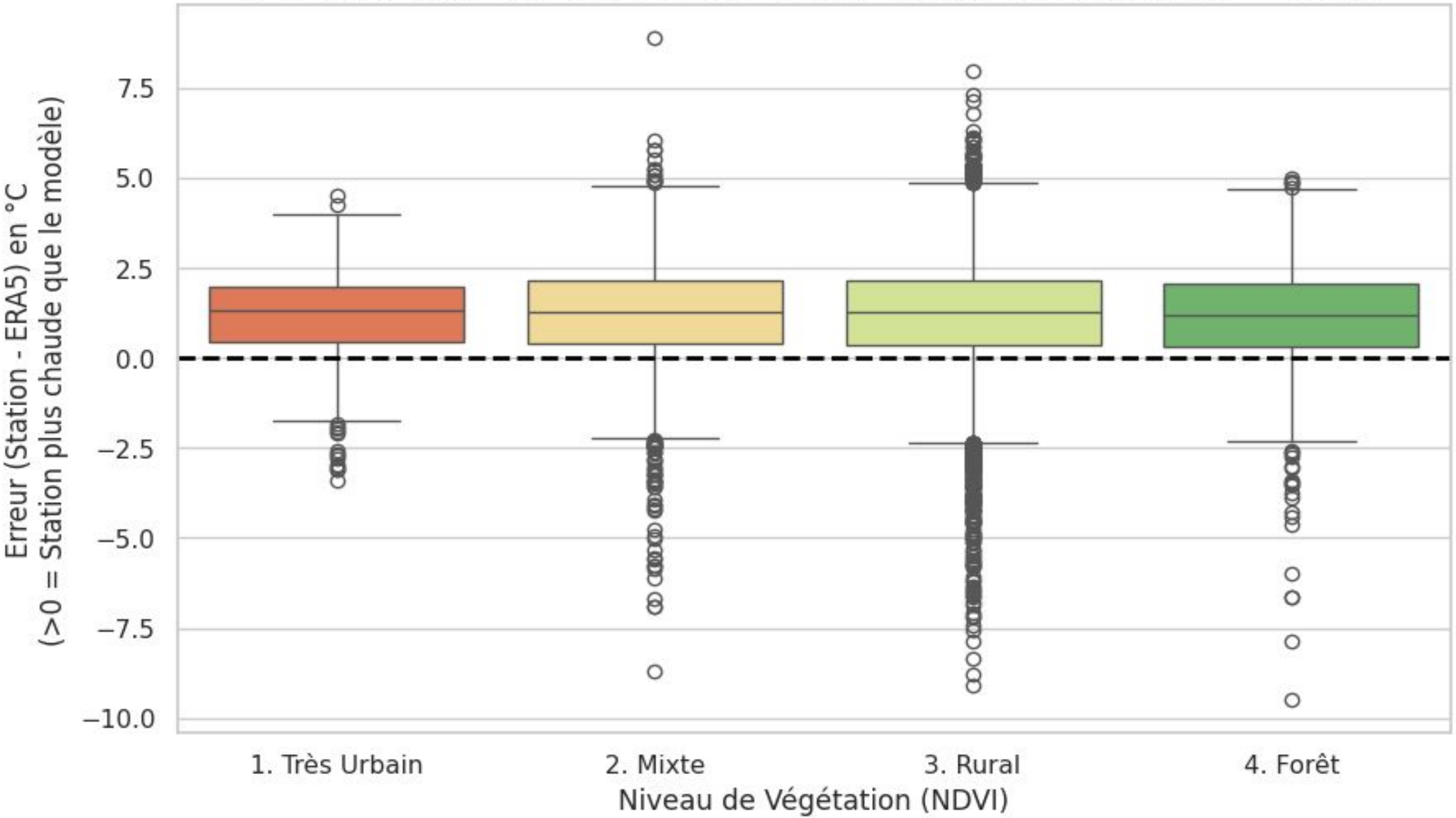
- **ERA5 underestimates temperature in low-NDVI (urban-like) environments**:
  mean bias = –1.04 °C.
- **ERA5 overestimates temperature in high-NDVI (vegetated) regions**:
  mean bias = +1.66 °C.
- **Error variability is highest in low-NDVI areas** (std = 3.57 °C), reflecting the strong heterogeneity of urban microclimates that ERA5 cannot resolve.
- **These patterns are consistent with an urban heat island signature**: ERA5 smooths strong urban warming and fails to capture localized heat amplification.



NDVI class	Mean bias	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

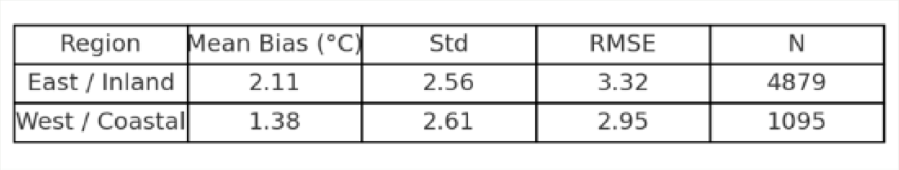


Note: Analysis performed on a curated subset of stations with complete daily data and valid NDVI sampling to ensure comparability across NDVI, season, and altitude classes.

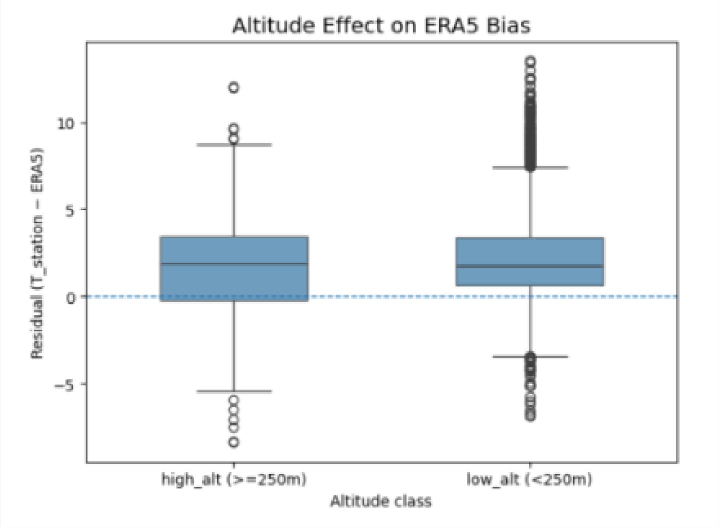

Erreur (Station - ERA5) en °C
(>0 = Station plus chaude que le modèle)
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
1. Très Urbain
2. Mixte
3. Rural
4. Forêt
Niveau de Végétation (NDVI)

Geographic Factors: Altitude, Latitude, Coastline

Geographic Factors: Altitude, Latitude, Coastline



Region	Mean Bias (°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 m): mean bias = +2.29 °C**
- **High-altitude (≥250 m): mean bias = +1.66 °C**

Low-altitude stations tend to be more **urban and inland**, where ERA5 struggles the most to capture local heat amplification.

Coastline effect

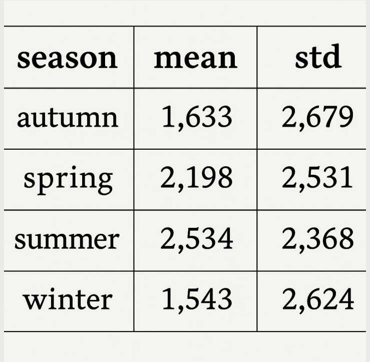
ERA5 underestimation is **stronger inland** (mean = +2.11 °C) than in west/coastal regions (mean = +1.38 °C). Coastal thermal gradients make ERA5 slightly more reliable there.

Seasonal Bias Patterns

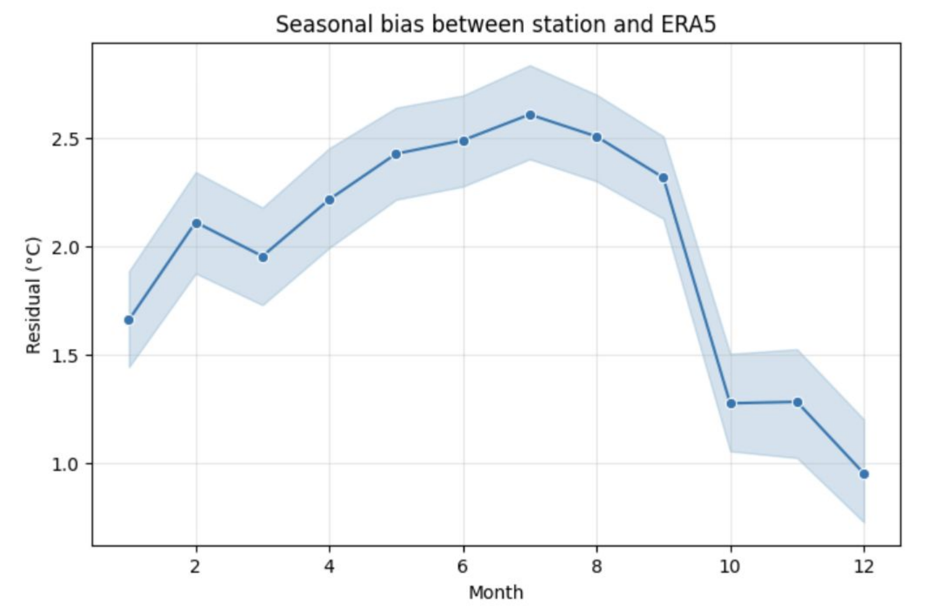
Seasonal Patterns Reveal Missing Urban Warming in Summer

The discrepancy between station data and ERA5 is strongly seasonal:

- ERA5 underestimates temperature year-round.
- Largest bias in summer (+2.53 °C) → strongest UHI + solar forcing.
- Moderate bias in spring (+2.20 °C).
- Lower bias in autumn (+1.63 °C) and winter (+1.54 °C).



season	mean	std
autumn	1,633	2,679
spring	2,198	2,531
summer	2,534	2,368
winter	1,543	2,624



Seasonal pattern matches known UHI dynamics: warm-season heat storage is not captured by coarse-grid reanalysis.

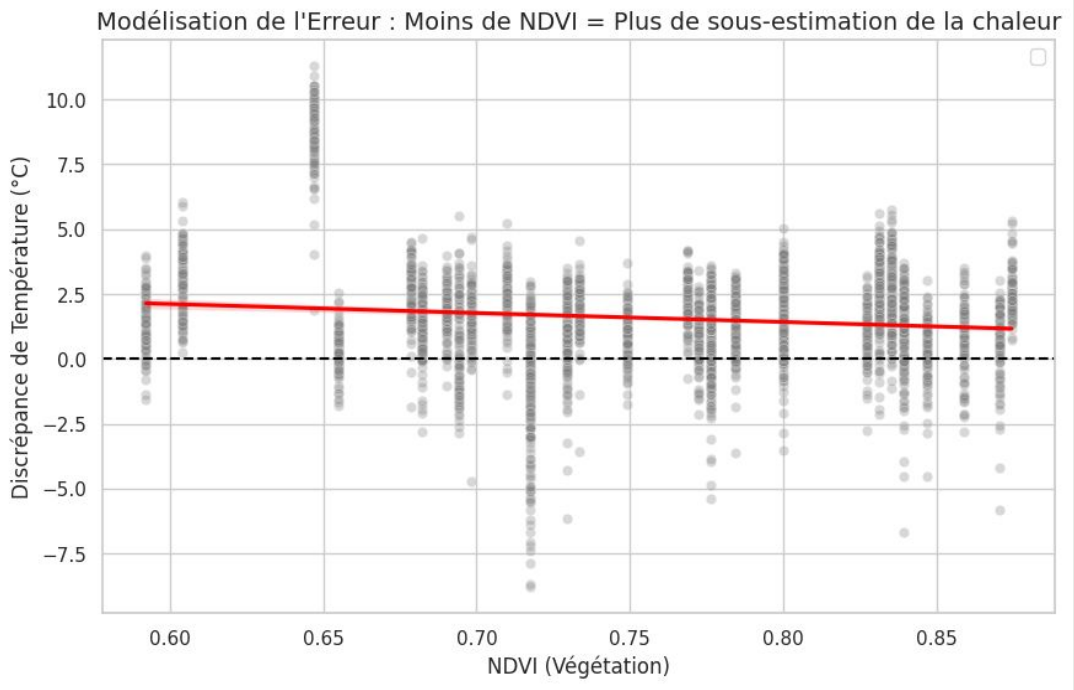
Impact of Vegetation on Temperature Prediction Error

Observations:

- Slight negative trend: higher NDVI → smaller error.
- Low vegetation areas (low NDVI): model often underestimates heat.
- High dispersion, especially around NDVI ≈ 0.65–0.70.

Interpretation:

- Densely vegetated areas: errors close to 0°C.
- Low vegetation / urban areas: temperature underestimated.
- Suggestion: account for vegetation to improve the model.
-


Modélisation de l'Erreur : Moins de NDVI = Plus de sous-estimation de la chaleur
Discrépance de Température (°C)
NDVI (Végétation)