GenHack 2025 - Week 3: Quantitative Metrics & Discrepancy Analysis

Urban Heat Island Analysis

Berlin, Germany

Team 19 UrbanCoolers

November 24, 2025

Week 3: Quantitative Metrics & Discrepancy Analysis

Unpacking the Systematic Biases in ERA5

Objective

To quantify and explain the discrepancies between ERA5 reanalysis data and ground station observations.

Scope

5 European cities, 286 weather stations, 4-year period (2020-2023)

Key Question

Where, when, and why does ERA5 fail to accurately represent surface temperatures?

ERA5 Shows a Consistent Cold Bias

−1.25°C

Mean Bias

ERA5 is consistently colder than observations

2.05°C

Overall Accuracy (RMSE)

0.99

Correlation

Excellent temporal co-variability

Key Takeaway: While ERA5 perfectly captures the *timing* of temperature changes (high correlation), it has a significant and systematic *cold bias* that must be corrected for reliable applications.

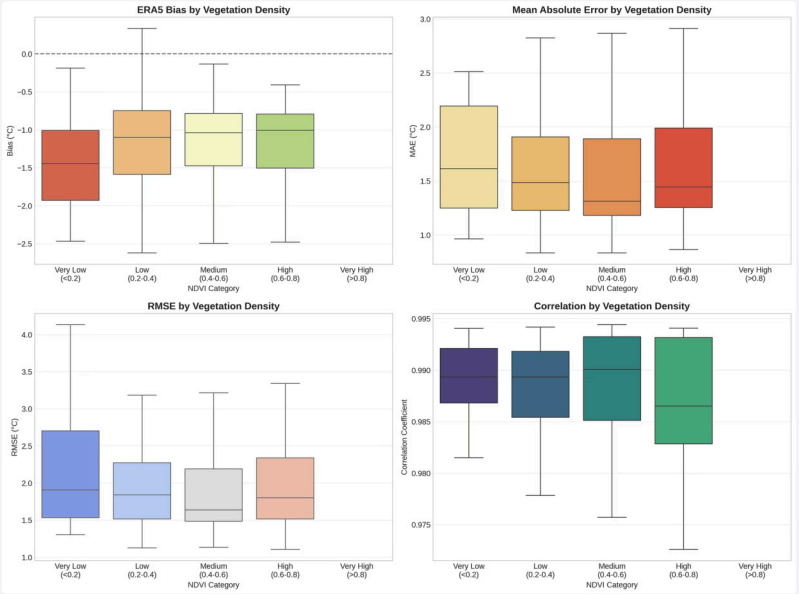
Metric 1: Error vs. Vegetation (NDVI)

Finding: No statistic ally significant relationship between vegetation density and ERA5 bias.

- **ANOVA p-value:** 0.37 (Not Significant)
- **Correlation (r):** 0.04 (Very weak)

NDVI Category	Mean Bias (°C)
Very Low	-1.63
Low	-1.13
Medium	-1.27
High	-1.24

Conclusion: The model's error is not primarily driven by how well it represents surface vegetation.



Metric 2: Error vs. Urbanization

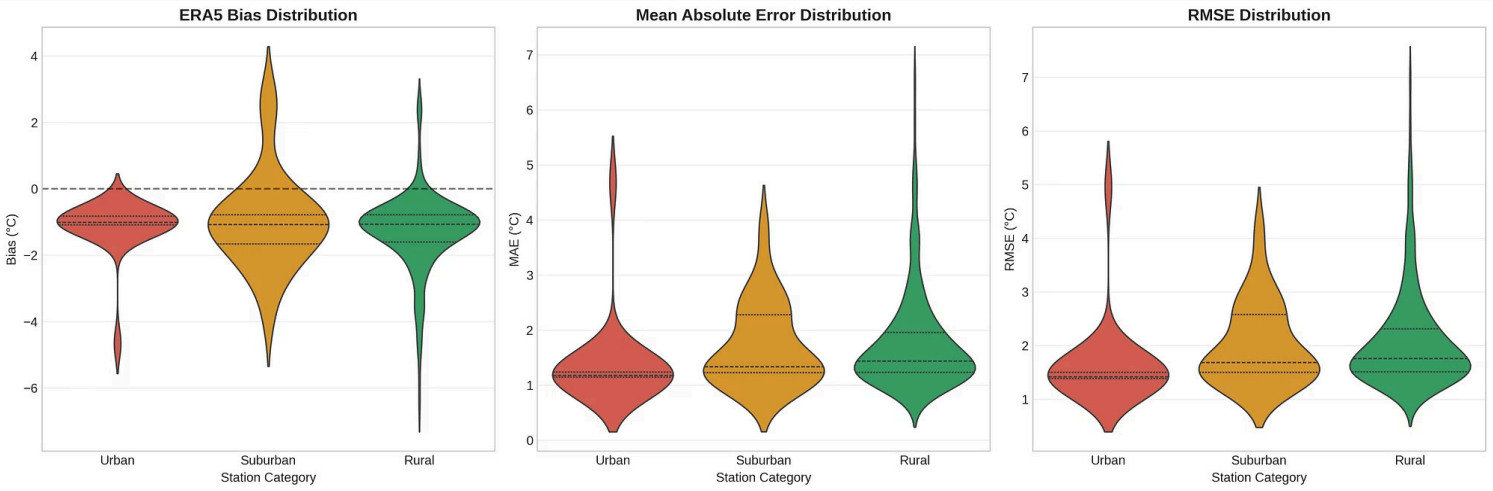
Finding: No statistically significant difference in bias between urban, suburban, and rural stations.

Category	Mean Bias (°C)
Urban	-1.14
Suburban	-0.97
Rural	-1.29

Statistical Test (Urban vs. Rural):

- **p-value:** 0.57 (Not Significant)
- **Mean Difference:** +0.15°C

Conclusion: ERA5's cold bias is pervasive across all landscapes; it is not a uniquely "urban" problem.



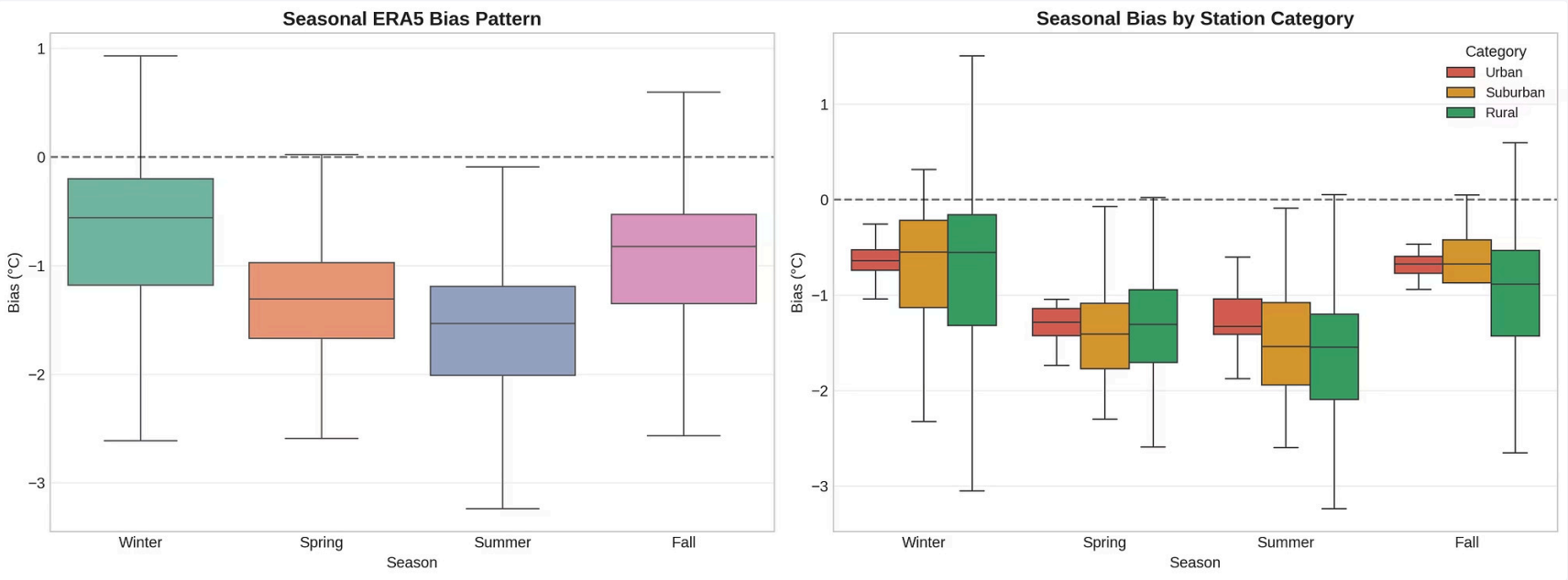
Metrics 3 & 4: Seasonal & Geographic Variation

Seasonal Pattern (Significant, $p < 0.001$):

- **Summer:** Largest cold bias (**-1.69°C**)
- **Winter:** Smallest cold bias (**-0.78°C**)

Geographic Pattern (Significant, $p < 0.01$):

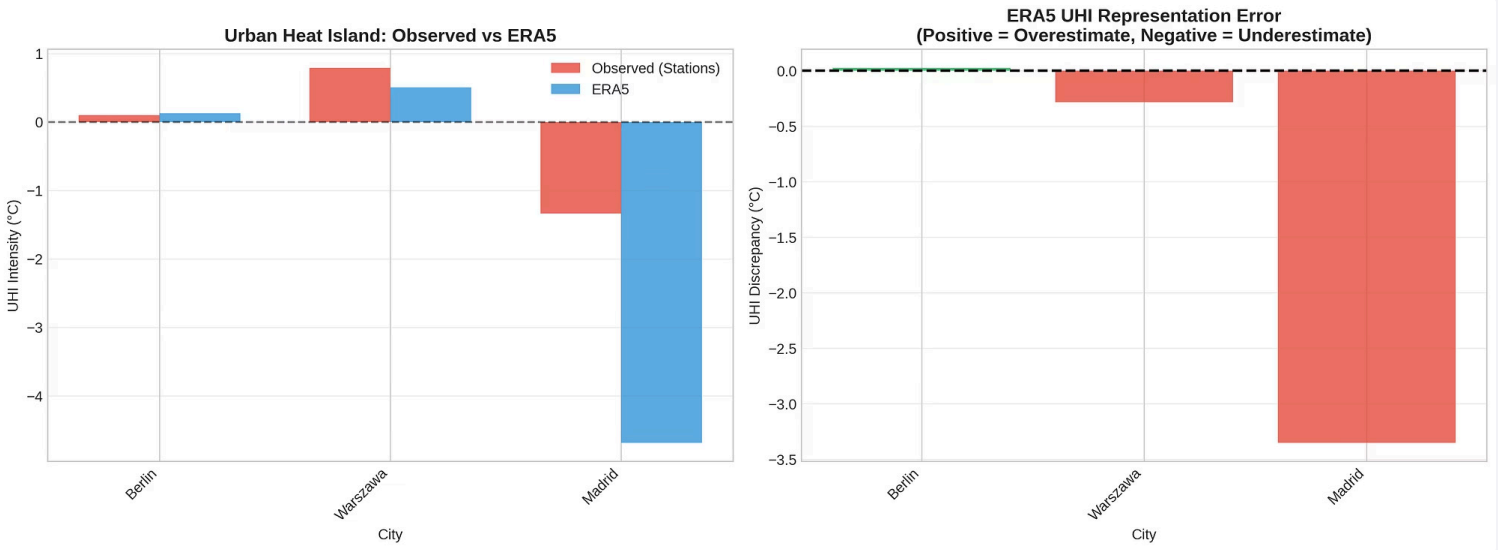
- **Milano:** Highest bias (-1.59°C)
- **Warszawa:** Lowest bias (-0.81°C)



Interpretation: The bias is **not constant**. It is strongest in summer and in certain topographically complex regions (e.g., Milano's Po Valley).

Metric 5: UHI Representation Quality

Finding: ERA5 systematically underestimates the Urban Heat Island effect.



City	Observed UHI	ERA5 UHI	ERA5 Discrepancy
Berlin	+0.10°C	+0.13°C	**+0.03°C** (Slight Overestimation)
Warszawa	+0.79°C	+0.51°C	**-0.28°C** (Underestimation)
Madrid	-1.34°C	-4.69°C	**-3.35°C** (Major Underestimation)

Conclusion: ERA5's coarse resolution smears out intense urban warming, a critical flaw for urban climate studies.

Statistical Modeling: Identifying Explanatory Variables

Random Forest Model (R^2 = 0.24) identified the top predictors of ERA5 bias:

01

Elevation (29.7%)

The single most important factor.

02

Mean Station Temp (23.3%)

Indicates a multiplicative relationship.

03

Latitude (16.8%)

Points to broad geographic trends.

04

Distance to Coast (8.6%)

05

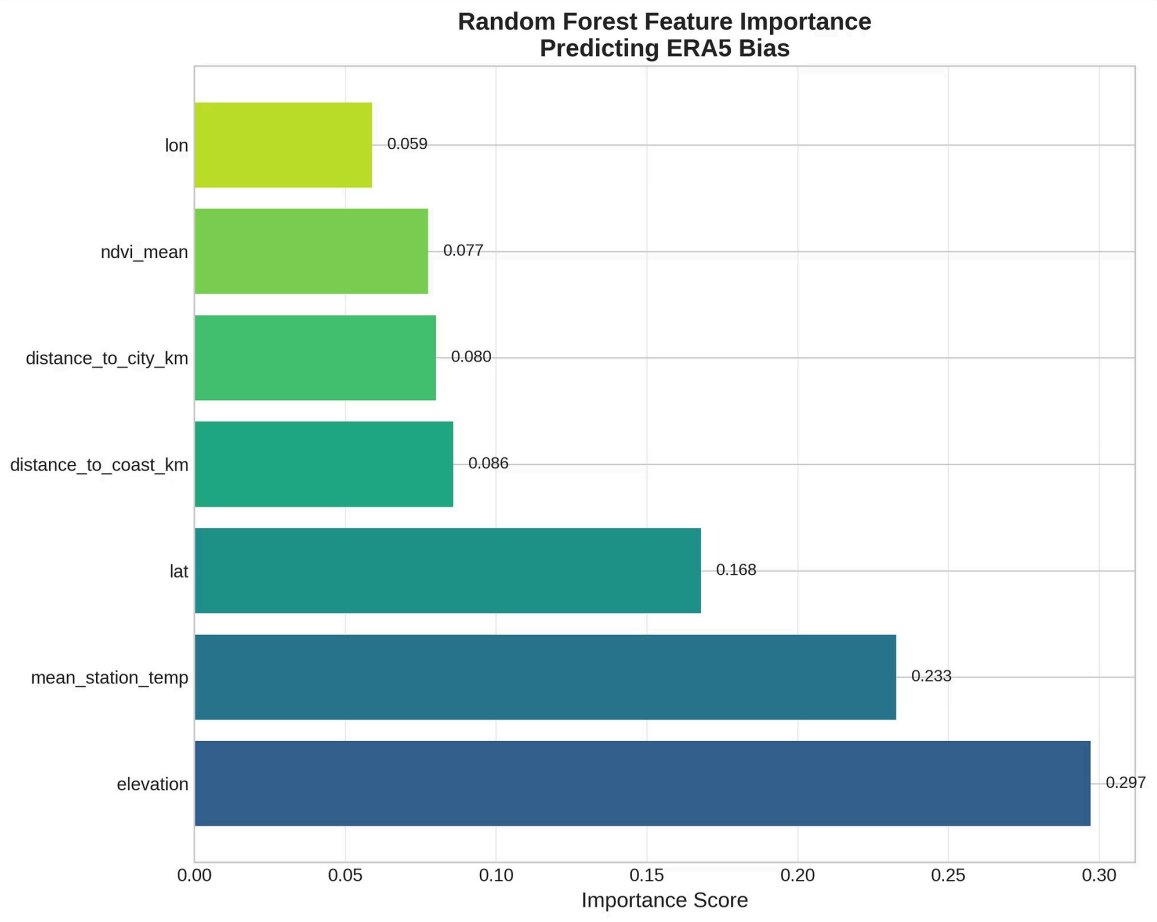
Distance to City (8.0%)

06

NDVI (7.8%)

Confirms its low importance.

Statistical Modeling: Identifying Explanatory Variables



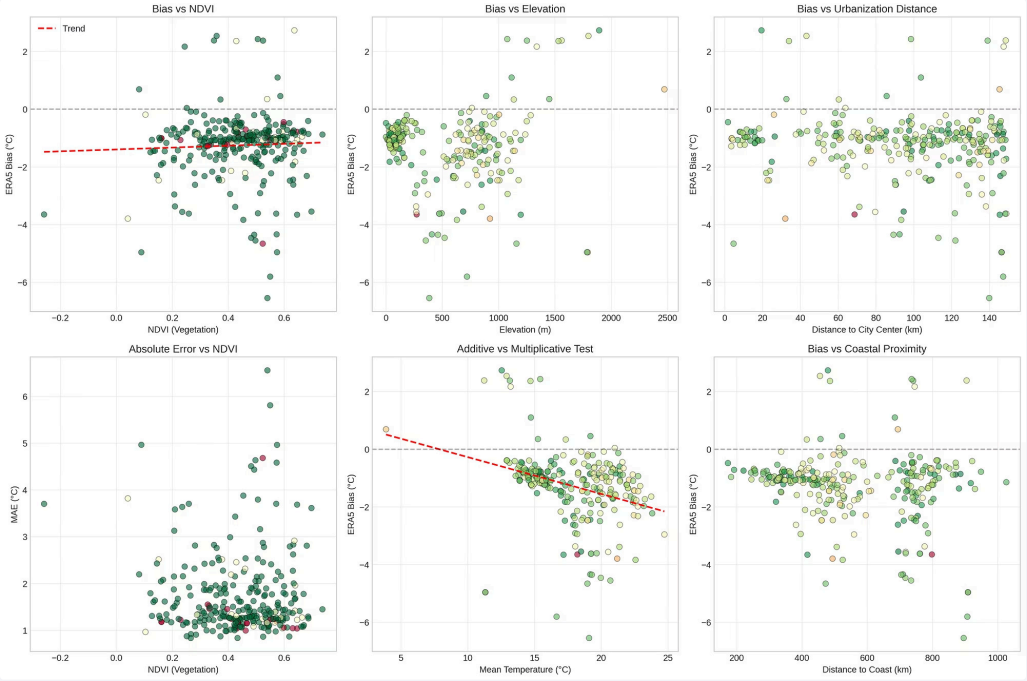
Key Insight: Topography and absolute temperature are the dominant controllers of ERA5 error, not land surface characteristics like vegetation or urbanization.

The Bias Scales with Temperature

Finding: A clear multiplicative component was detected.

Correlation (Bias vs. Temp)

r = -0.33 ($p < 0.001$)



Interpretation:

- The cold bias is **worse at lower temperatures** and **less severe at higher temperatures**.
- This is an **unexpected inverse relationship** that contradicts the simple assumption that models struggle more with extreme heat.
- Suggests the bias correction may need to be temperature-dependent.

Summary & Next Steps

1 Systemic Cold Bias

ERA5 is consistently ~1.25°C too cold in this region.

2 Top-Down Drivers

Error is best explained by **elevation and background climate**, not local surface details (NDVI/Urbanization).

3 UHI Blind Spot

The model significantly **underrepresents urban heating**.

4 Complex Dynamics

The bias is **seasonal and multiplicative**, requiring a non-constant correction.

Implications for Week 4 (Bias Correction):

- A simple additive correction will be insufficient.
- The correction model **must incorporate elevation and temperature**.
- Special attention is needed to correct the **UHI representation** in urban areas.