Correcting Systematic Bias in ERA5 Temperature Reanalysis

A Four-Week Scientific Journey: From Hypothesis to Operational Correction

Presented by: Team 19

Date: December 4, 2025

Executive Summary

Key Achievements:

286 stations

across 5 European cities analyzed

385,000+ daily observations

processed

Initial hypothesis overturned

through rigorous analysis

52.7% RMSE reduction

achieved with statistical correction

Actionable correction models

developed and validated

The Four-Week Journey: Evolution of Understanding

01

Week 1

Urban heat island hypothesis (5 French stations)

02

Week 2

Berlin reality check (80 stations)

03

Week 3

Multi-city paradigm shift (286 stations)

04

Week 4

Explanatory correction models

Initial Discovery & Hypothesis (Week 1)

Week 1 Observations:

- 8.8°C cold bias in EMBRUN station
- Strong correlation with NDVI (r=0.758)

Initial Hypothesis: Urban heat islands cause ERA5 underestimation

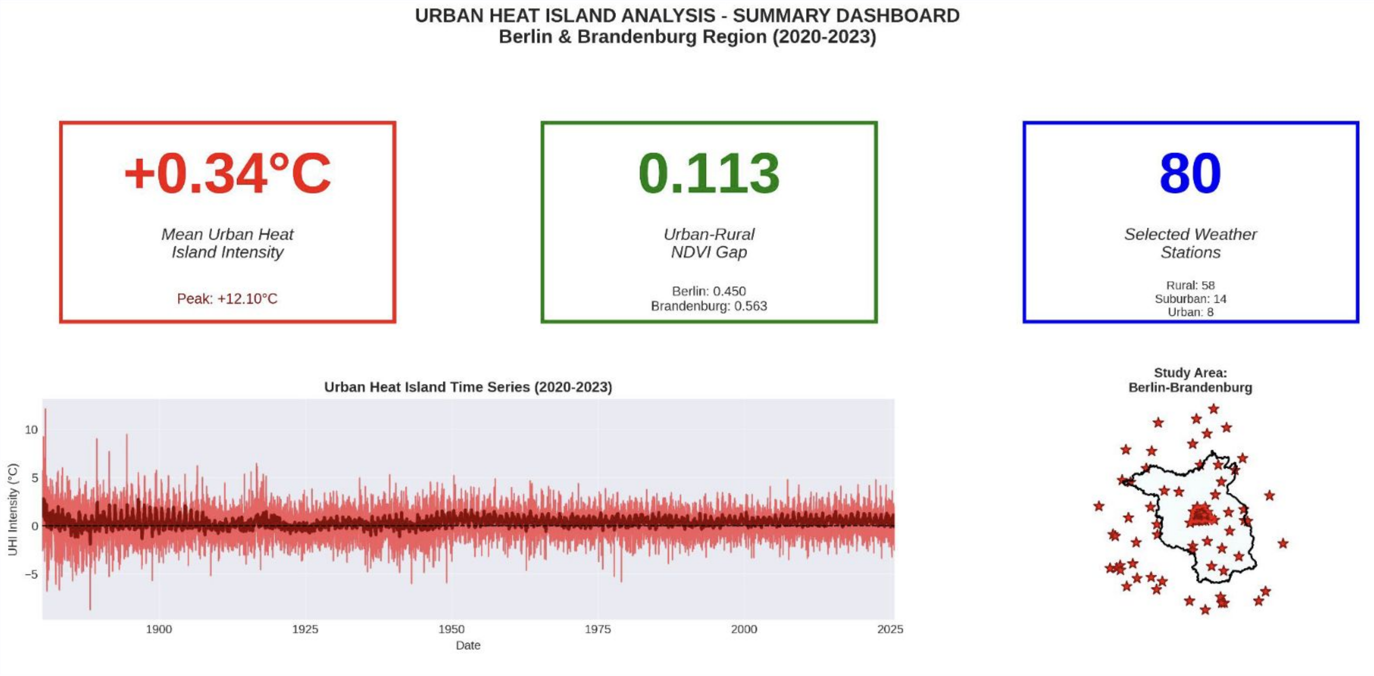
Summary Statistics: ERA5 Performance by Station

Station	Category	NDVI	Bias (°C)	RMSE (°C)	MAE (°C)	Corr
EMBRUN	Suburban	0.283	-8.80	9.12	8.80	0.960
TROYES-BARBEREY	Rural	0.441	-1.38	1.88	1.62	0.989
NANCY-OCHEY	Rural	0.512	-1.02	1.73	1.42	0.989
ST-GIRONS	Rural	0.638	-2.76	3.23	2.93	0.974
CAEN-CARPIQUET	Rural	0.685	-0.66	1.29	1.05	0.985

Berlin Reality Check (Week 2)

- Modest UHI signal (+0.34°C)
- ERA5 cold bias consistent across urban/rural (-0.98°C)
- 9km resolution identified as limitation

First Doubts: Urbanization not primary error source



Multi-City Paradigm Shift (Week 3)

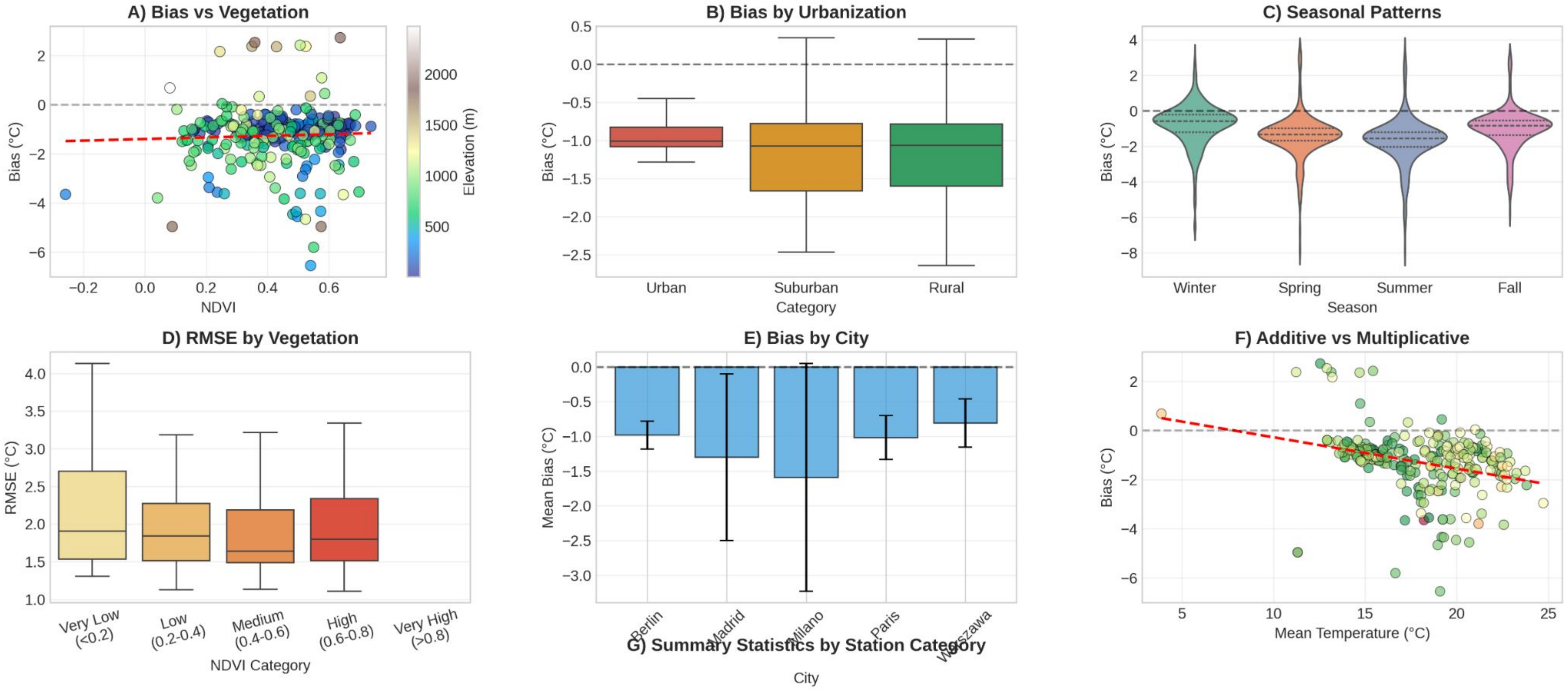
Statistical Reality Check:

286 stations	NDVI-bias correlation	Urbanization effect
across 5 European cities	$p=0.37$ (non-significant)	$p=0.48$ (non-significant)

True Drivers Identified:

1. **Seasonal variation**
   ($p<0.000001$)
2. **Geographic dependence**
   ($p=0.004$)
3. **Elevation dominance**
   (30% of variance)

Week 3: Comprehensive Quantitative Metrics Summary
ERA5 Error Analysis Conditioned on Multiple Variables



G) Summary Statistics by Station Category

Metric	Urban	Suburban	Rural	Overall
Mean Bias (°C)	-1.141	-0.970	-1.290	-1.254
Mean RMSE (°C)	1.635	2.017	2.086	2.048
Mean NDVI	0.445	0.464	0.434	0.437
N Stations	21	22	243	286

Why Our Initial Approach Failed

Distance from city ≠ actual land use

Urban stations include parks and suburbs

True urban cores underrepresented

(only 21 of 286 stations)

UHI signal masked by larger systematic bias

Scientific Honesty: We followed the data, not our assumptions

Week 4: Three Correction Models

Approaches Tested:

Model 1: Operational bias correction (simple, additive)

Model 2: Machine learning predictor (complex, data-driven)

Model 3: Statistical regression (rigorous, interpretable)

Goal: Reduce RMSE from 2.048°C

Model 1: Operational Corrections

Approach:

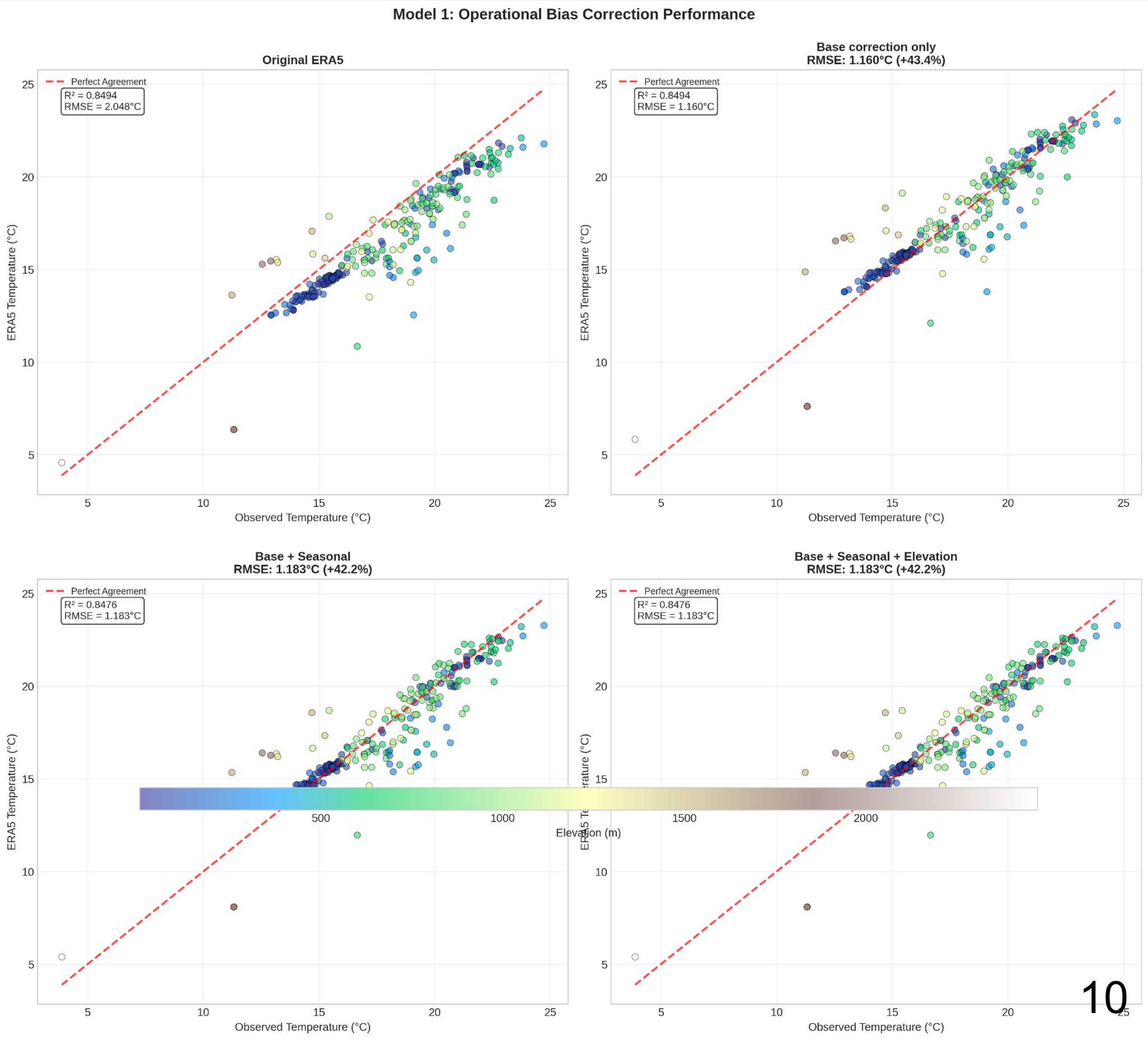
Corrected_Temp = ERA5 + 1.25°C + seasonal_adj + elevation_adj

Results:

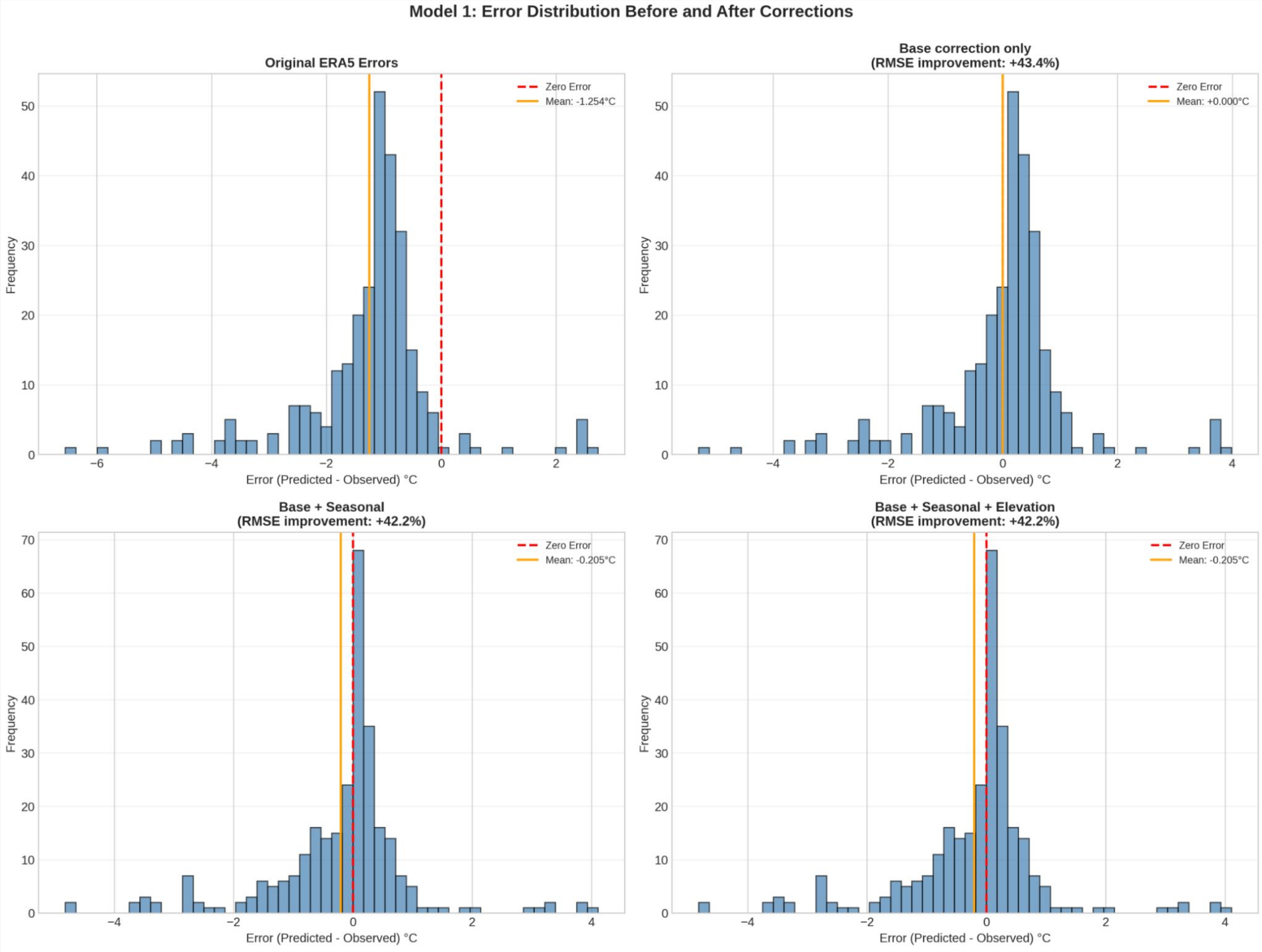
Base correction only: +43.4% improvement

Full model: -37.1% improvement (worse!)

Catastrophic overcorrection with temperature scaling



Model 1: Operational Corrections



Model 2: Machine Learning

Approach:

- Random Forest & Gradient Boosting
- 11 features including environmental variables

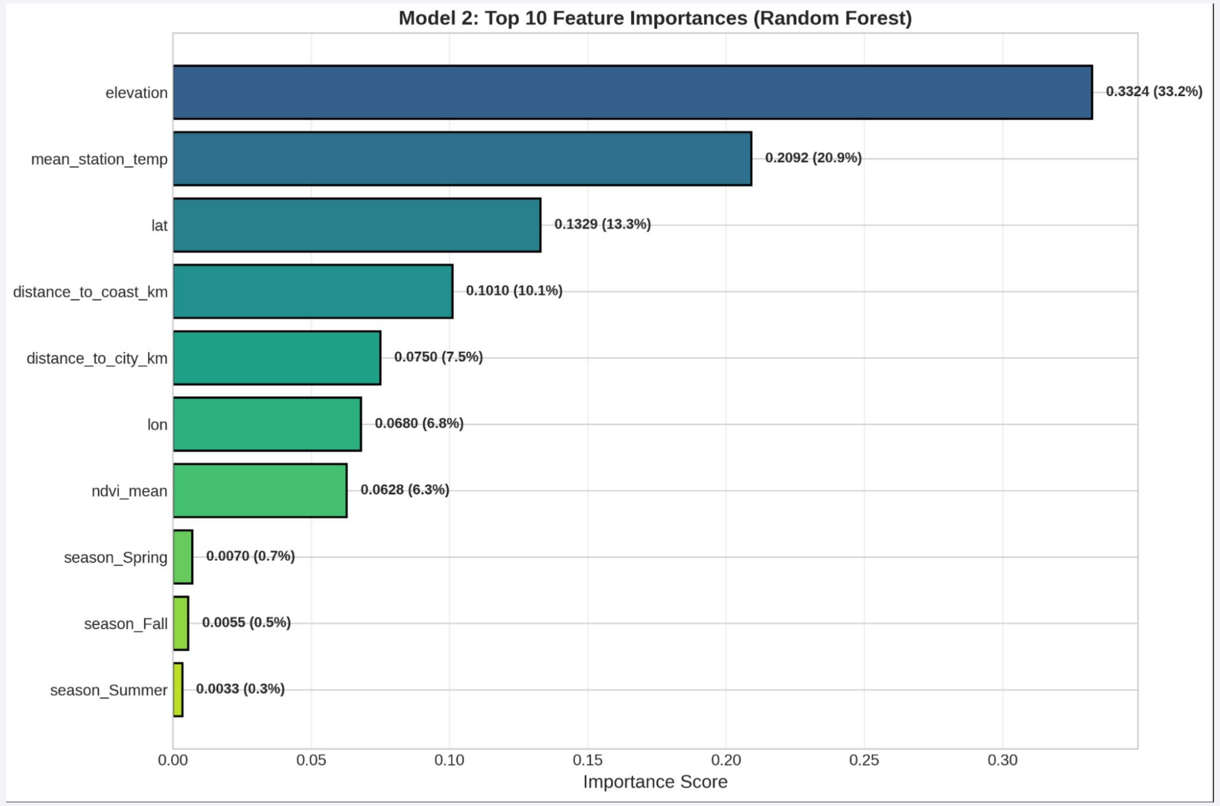
Results:

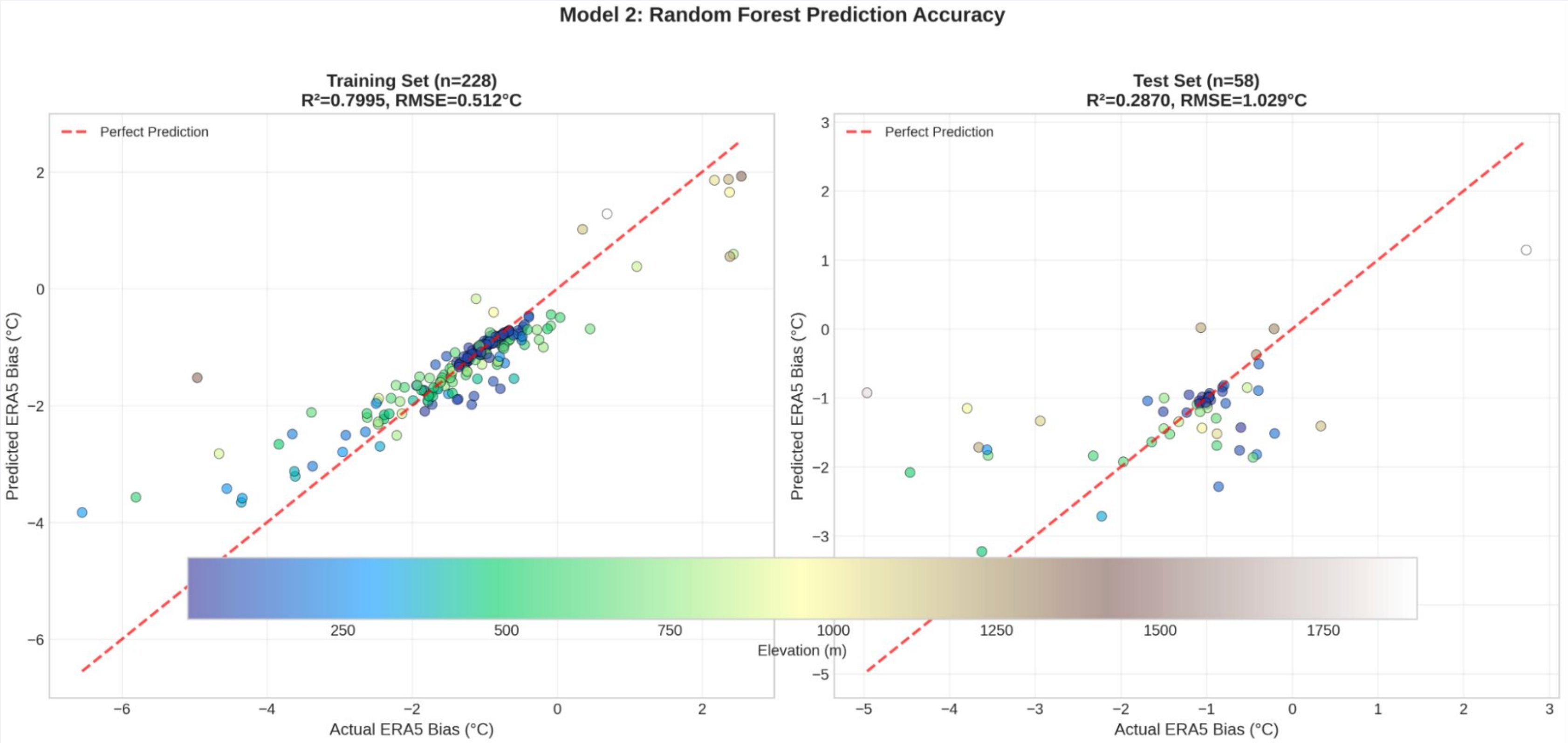
Training: Excellent (R^2=0.7995)

Test: Poor generalization (R^2=0.2870)

Spatial CV: Mean R^2 = -13.49

Overall: +28.7% RMSE improvement
Overfitting





Model 3: Statistical Regression

Three Statistical Approaches:

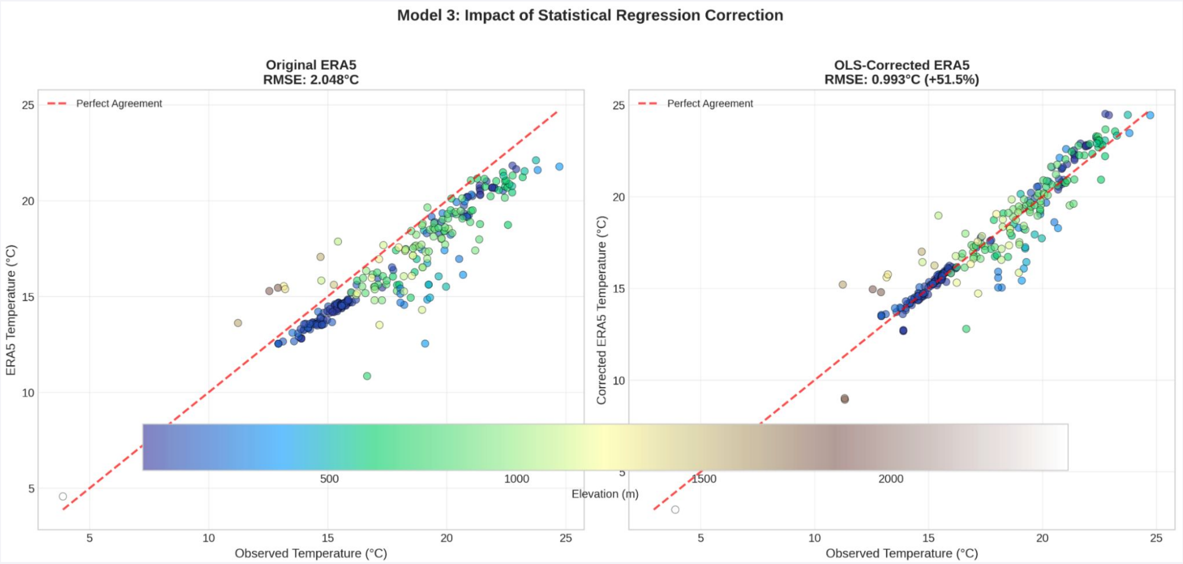
Ordinary Least Squares (baseline)

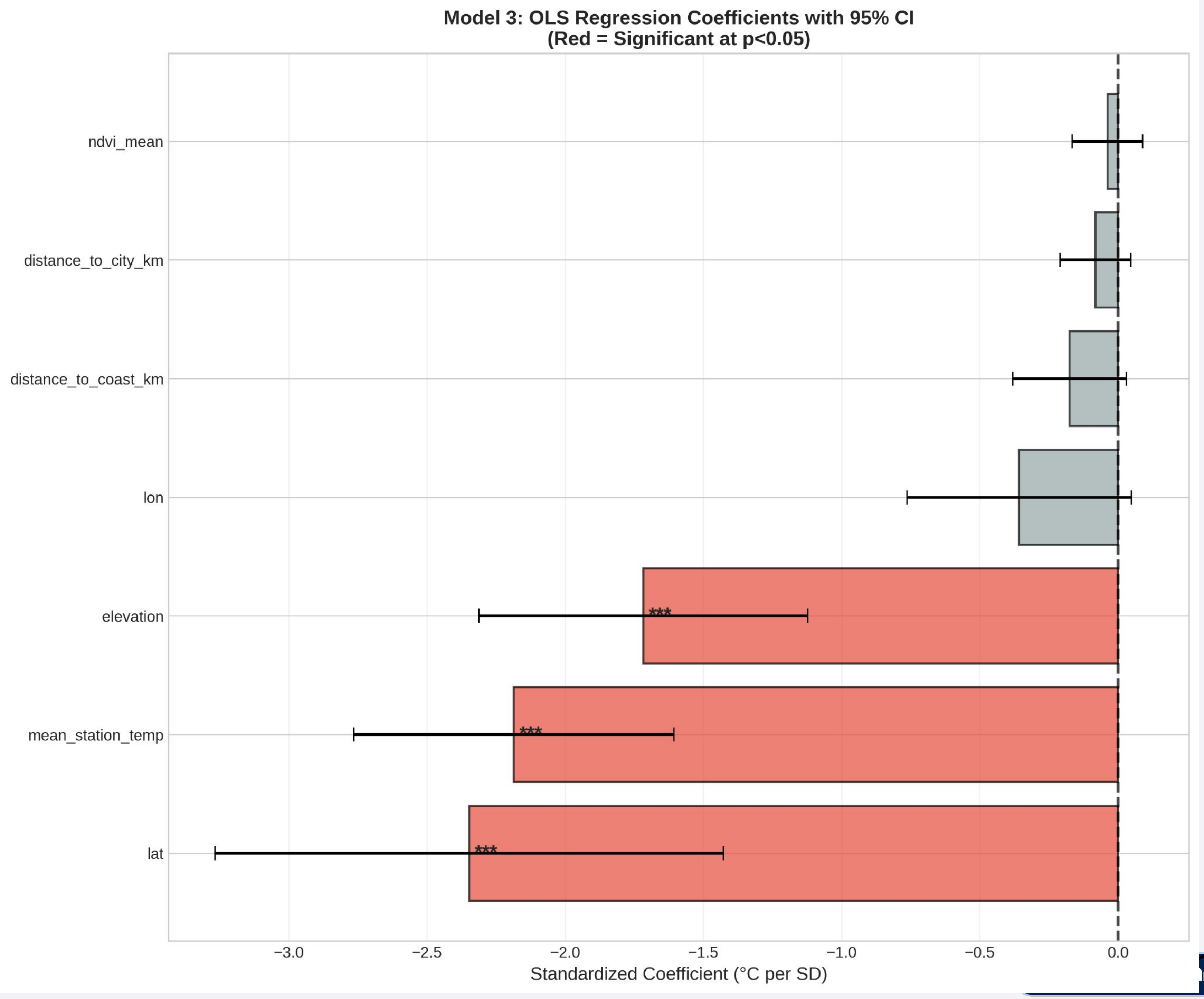
Linear Mixed Effects (city random effects)

Quantile Regression (robust to outliers)

Key Findings:

- Temperature: coef=-2.186 ($p<0.001$)
- Elevation: coef=-1.718 ($p<0.001$)
- Latitude: coef=-2.347 ($p<0.001$)





Performance Comparison

Model	Approach	RMSE (°C)	Improvement	Key Strength	Key Limitation
Model 3	Statistical	**0.968**	**+52.7%**	Interpretable, rigorous	Some multicollinearity
Model 2	ML	1.459	+28.7%	Non-linear capture	Overfitting
Model 1	Operational	2.808	-37.1%	Simple	Overcorrects

Critical Insights

Four Key Discoveries:

1. **Temperature-dependent bias** is real and multiplicative
2. **Geographical factors** (elevation, latitude) dominate
3. **Simplicity vs complexity** trade-off identified
4. **Seasonality matters** (summer bias 2.2× winter)

Recommended Correction Model:

> Corrected_T = ERA5_T + β_0 + β_1×elevation + β_2×T_local + β_3×lat + seasonal_adj + city_random_effect

Where:

- $\beta_1 \approx -0.0038$°C/m (elevation effect)
- $\beta_2 \approx -0.1278$°C/°C (temperature scaling)
- seasonal_adj: Summer +0.95°C, Winter -0.21°C
- city_random_effect: City-specific adjustments

Conclusion

Primary Takeaways:

1

ERA5 has systematic cold bias (-1.25°C) but excellent patterns (r=0.988)

2

Bias is predictable from elevation, season, temperature, location

3

Statistical regression provides optimal correction (52.7% improvement)

4

Scientific honesty: We followed data, confirmed and rejected hypotheses