Contents

1 Methodology

1.1 Overview

This script evaluates the potential cooling effects of retrofitting existing roofs with green roofs by combining satellite-derived biophysical indicators with a machine-learning based empirical Land Surface Temperature (LST) model. The workflow is inspired by recent remote sensing and urban heat island (UHI) literature, including the Random Forest models developed for Belgian cities (Joshi et al., 2023), the approach used in Granada (Sánchez-Cordero et al., 2025), causal inference work in Durham (Calhoun et al., 2024), and machine learning LST modelling studies such as the Yazd case study (Mansourmoghaddam et al., 2024).

The methodology consists of six major stages:

1. Acquisition and preprocessing of Landsat 8/9 Collection 2 Level-2 products.
2. Computation of NDVI, broadband albedo, and NDBI from Landsat Surface Reflectance (SR) bands.
3. Extraction of surface temperature from Landsat Surface Temperature (ST) products.
4. Spatially-thinned sampling of predictors ($\geq$100 m spacing) to reduce spatial autocorrelation.
5. Training of a Random Forest LST model using the SR-derived predictors.
6. Simulation of a green roof scenario by modifying rooftop NDVI and albedo, followed by LST prediction and computation of ΔLST.

1.2 Satellite Data

1.2.1 Landsat 8/9 Surface Reflectance (SR)

Landsat 8/9 Collection 2 Level 2 SR imagery provides atmospherically-corrected reflectance in spectral bands used to derive key UHI predictors. Three biophysical variables were computed:

- Normalized Difference Vegetation Index (NDVI) …
- Broadband albedo, based on the full Liang formula (Liang, 2001; Liang et al., 2003):

$$\alpha = 0.356\rho_{\text{Blue}} + 0.130\rho_{\text{Red}} + 0.373\rho_{\text{NIR}} + 0.085\rho_{\text{SWIR1}} + 0.072\rho_{\text{SWIR2}} + 0.0018.$$

- Normalized Difference Built-up Index (NDBI) …

NDVI and NDBI are theoretically justified and widely used predictors of LST (Joshi et al., 2023; Mansourmoghaddam et al., 2024; Calhoun et al., 2024). Broadband albedo modifies radiative heat fluxes and is strongly related to surface temperature (Liang et al., 2003).

1.2.2 Landsat 8/9 Surface Temperature (ST)

The ST product (ST_B10) provides physically corrected LST derived from the TIRS sensor and corrected for emissivity and atmospheric effects. This product serves as the target variable for the empirical model.

1.3 Building Data and Roof Types

Building footprints with roof type attributes were used to identify roofs that can be converted to green roofs. All building geometries were reprojected to the Landsat CRS to ensure raster alignment. A supersampled rasterization procedure estimates, for each Landsat pixel, the fractional area covered by targeted roof types.

1.4 Computation of Predictor Variables

All predictor rasters (NDVI, albedo, NDBI) were computed at 30 m resolution, aligned to the ST raster grid. Invalid SR or ST pixels (e.g. water, cloud, shadow) were excluded via QA masks as is standard in Landsat-based UHI studies.

1.5 Spatial Sampling with Minimum 100 m Spacing

Following the recommendations of Joshi et al. (2023) and Sánchez-Cordero et al. (2025), spatial autocorrelation is reduced by enforcing an approximate minimum spacing of 100 m between training pixels.

Since Landsat pixels are 30 m, a grid-based thinning scheme was used:

1. The scene is divided into non-overlapping blocks of size $b \times b$ pixels, where:

$$b = \mathrm{round}\left(\frac{100\,\mathrm{m}}{30\,\mathrm{m/pixel}}\right) \approx 3\text{–}4.$$

2. Within each block, one valid pixel is selected at random.
3. From this set of spatially independent pixels, a random subset (typically 10%) is used for model training.

This method approximates Poisson-disk sampling and significantly reduces model overfitting caused by clustered training points.

1.6 Random Forest LST Model

A Random Forest regressor (Breiman, 2001) predicts LST from the three predictors:

$$\mathrm{LST} = f(\mathrm{NDVI}, \alpha, \mathrm{NDBI}),$$

where f is the ensemble of decision trees.

Random Forests were selected because urban surface temperature exhibits complex, non-linear responses to vegetation, built-up intensity, and radiative properties, as widely documented in recent empirical and machine-learning studies (Joshi et al., 2023; Mansourmoghaddam et al., 2024). In this context, Random Forests offer several advantages:

- they effectively capture the **non-linear relationships** between LST and predictors such as NDVI, albedo, and NDBI (Joshi et al., 2023; Mansourmoghaddam et al., 2024),
- they are robust to noise, outliers, and multicollinearity among predictor variables,
- and they consistently outperform linear models in LST prediction across a broad range of urban heat studies.

An 80/20 train-test split is used. Because spatial thinning reduces autocorrelation, a conventional random split is acceptable for single-scene analysis (as argued in (Joshi et al., 2023)).

Typical model performance for high-quality summer scenes reached:

$$R^2_{\text{test}} \approx 0.60\text{–}0.75, \quad \text{RMSE} \approx 2.0\text{–}2.3^\circ\text{C},$$

consistent with existing RF-based LST studies.

1.7 Green Roof Scenario Simulation

For all targeted roof types, NDVI and albedo were modified to values representing typical extensive green roofs:

$$\text{NDVI}_{\text{green}} = 0.4, \qquad \alpha_{\text{green}} = 0.20.$$

These values follow empirical observations documented in (Sánchez-Cordero et al., 2025; McConnell et al., 2022; Yan et al., 2025) and general green roof vegetation literature.

For any pixel with roof fraction f, predictor modification is applied only proportionally:

$$\text{NDVI}' = (1 - f) \cdot \text{NDVI} + f \cdot 0.4,$$
$$\alpha' = (1 - f) \cdot \alpha + f \cdot 0.20.$$

NDBI remains unchanged, as it represents structural built-up characteristics.

The Random Forest model is then applied to the modified predictors to produce the green roof scenario LST:

$$\mathrm{LST}' = f(\mathrm{NDVI}', \alpha', \mathrm{NDBI}).$$

Pixels without modified roofs retain the original Landsat LST.

The cooling effect is computed as:

$$\Delta\mathrm{LST} = \mathrm{LST}' - \mathrm{LST}.$$

Building-level results are obtained using zonal statistics.

1.8 Summary

This methodology follows recent scientific best practices in remote sensing UHI modelling:

- Landsat SR + ST as primary data sources,
- NDVI, NDBI, and broadband albedo as predictors,
- spatially thinned sampling to reduce spatial autocorrelation,
- Random Forest regression for empirical LST modelling,
- scenario-based NDVI/albedo modification for green roof simulation.

The approach is fully reproducible, computationally efficient, and aligned with current UHI mitigation research.

Bibliography

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Calhoun, Z. D., Willard, F., Ge, C., Rodriguez, C., Bergin, M., and Carlson, D. (2024). Estimating the effects of vegetation and increased albedo on the urban heat island effect with spatial causal inference. *Scientific Reports*, 14(1):540.

Joshi, M. Y., Aliaga, D. G., and Teller, J. (2023). Predicting urban heat island mitigation with random forest regression in belgian cities. In Goodspeed, R., Sengupta, R., Kyttä, M., and Pettit, C., editors, *Intelligence for Future Cities: Planning Through Big Data and Urban Analytics*, The Urban Book Series, pages 305–323. Springer, Cham.

Liang, S. (2001). Narrowband to broadband conversions of land surface albedo i: Algorithms. *Remote Sensing of Environment*, 76(2):213–238.

Liang, S., Shuey, C. J., Russ, A. L., Fang, H., Chen, M., Walthall, C. L., Daughtry, C. S. T., and Hunt, R. (2003). Narrowband to broadband conversions of land surface albedo: Ii. validation. *Remote Sensing of Environment*, 84(1):25–41.

Mansourmoghaddam, M., Rousta, I., Ghafarian Malamiri, H., Sadeghnejad, M., Krzyszczak, J., and Ferreira, C. S. S. (2024). Modeling and estimating the land surface temperature (lst) using remote sensing and machine learning: Case study yazd, iran. *Remote Sensing*, 16(3):454.

McConnell, K., Braneon, C., Glenn, E., Stamler, N., Mallen, E., John, D. P., Pandya, R., Abramowitz, J., Fernandez, G., and Rosenzweig, C. (2022). A quasi-experimental approach for evaluating the heat mitigation effects of green roofs in chicago, illinois. *Sustainable Cities and Society*, 76:103376.

Sánchez-Cordero, F., Nanía, L., Hidalgo-García, D., and López-Chacón, S. R. (2025). Assessing the spatial benefits of green roofs to mitigate urban heat island effects in a semi-arid city: A case study in granada, spain. *Remote Sensing*, 17(12):2073.

Yan, X., Liu, J., Zhang, W., Wang, Z., Jia, R., Wang, L., Liu, N., and Zhu, Y. (2025). Examination of green roofs' cooling effects in a big city based on multi-source remote sensing data: An example from central xi'an, china. *Sustainable Cities and Society*, 130:106545.