Politecnico di Torino

Planning for the Global Urban Agenda

INTEGRATED MODELING OF THE SURFACE URBAN HEAT ISLAND USING LANDSAT REMOTE SENSING AND GIS TOOLS

A Comparative Assessment of Wrocław and Turin

A.a. 2025 / 2026

Master of Science July 2026

Relatrice / Relatore	Candidato / Candidata
Prof.ssa Guglielmina Mutani	Rouzbeh Ebrazi

Abstract

This thesis aims to investigate the spatial drivers of Surface Urban Heat Island Intensity (SUHII) in two European case studies: Wrocław, Poland and Turin, Italy. The objective is to map surface thermal anomalies and to check how far satellite-derived indicators, urban morphology variables and GIS-based modeling can explain their spatial distribution. The workflow was first developed for Wrocław and then applied to Turin following the same main procedure so that the two cities could be compared with consistent data processing steps. Landsat 8/9 Collection 2 Level 2 imagery was processed using Google Earth Engine to derive Land Surface Temperature (LST), SUHI, SUHII, UTFVI, and several surface indicators including NDVI, NDBI, NDWI, NDMI, Albedo, PVI and Emissivity. These raster outputs were then aggregated at the census section level in QGIS using zonal statistics. Additional morphology related variables such as Sky View Factor (SVF), Building Coverage Ratio (BCR) and Surface to Volume (S/V) Ratio were added to the final datasets. We processed the atmospheric conditions for each selected Landsat acquisition to describe the ERA5-Land meteorological data using the Python language. The analysis was structured in three interrelated phases. Multi-seasonal spatial assessment was first performed to compare LST, SUHI, SUHII, UTFVI and recurrent hotspot patterns among the selected seasons.

Second, Random Forest regression models were trained on the warm-season scenes, when the surface thermal contrast was the most pronounced. SUHII was used as the modeling target, while LST was excluded from the predictors to avoid target leakage. The final predictor set for both cities included Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio. Third, the spatial results were read together with a rule-based Local Climate Zone (LCZ) planning interpretation. LCZ was used only to support the planning discussion and was not included as a Random Forest predictor.

The Random Forest models produced good test-set results, with R^2 values of 0.7576 for Wrocław and 0.7841 for Turin. In Wrocław, the most important predictors were NDBI (0.3049), NDMI (0.2359), PVI (0.1029), Emissivity (0.0823), and BCR (0.0698). This indicates that built-up intensity and surface moisture conditions were central to the warm-season SUHII pattern. In Turin, the strongest predictors were Emissivity (0.1759), PVI (0.1640), NDWI (0.1489), NDVI (0.1449), NDBI (0.1408), and NDMI (0.1323), showing a stronger role of radiative surface behavior together with vegetation and moisture-related variables.

Results show that the interaction of built-up surfaces, vegetation, moisture and surface radiative properties affects both cities but the relative importance of these factors varies depending on the local context. In Wrocław, the spatial evidence supports planning actions of protecting blue-green continuity, increasing shading and improving surface permeability in thermally stressed areas. The results in Turin are more directly pointing towards a dense and sealed urban fabric, where cool or reflective materials, targeted greening, green roofs and permeability improvements can help to reduce surface overheating.

These interventions are proposed as planning recommendations, based on the thermal evidence mapped and model interpretations. They are not simulated as before-after regeneration scenarios in this thesis.

Acknowledgements

To Professor Guglielmina Mutani.

My deepest thanks to Prof. Guglielmina Mutani. Without her kindness, guidance and support, I wouldn't have been able to complete this academic path. Her guidance, patience and constructive comments were instrumental in shaping this thesis and allowing me to bring this work to completion.

To my parents.

My most sincere thanks to my parents for their patience and support. Even from afar their unconditional love, encouragement and support throughout my studies stood by me. It is because of them, above all, that I am standing here today, completing this stage of my academic journey.

To my fiancée.

I would also like to thank my fiancée sincerely, who was there for me during the development of this project. Her endless support, patience and encouragement through the tough and rewarding moments of this work gave me strength and motivation.

Dedication.

Finally, I would like to dedicate this thesis to the more than 36,000 defenseless people of Iran who were massacred while peacefully demanding freedom and fundamental rights. Their memory remains a lasting symbol of courage, dignity, resistance, and hope.

Table of contents

Contents

1. Literature Review

1.1 General overview of the analysis

Surface Urban Heat Islands (SUHIs) are an important topic in urban climate studies because they show how land cover, surface materials, vegetation, and urban form influence the thermal behavior of cities. This literature review introduces the main concepts needed for the thesis and explains how SUHI can be assessed through satellite remote sensing and spatial analysis. It focuses on the physical causes of surface overheating, the implications of the higher urban temperatures and the key mitigation approaches discussed in previous studies [1]. The literature review also helps to determine the methodological position of the work for this thesis. Landsat imagery is relevant in that it contains spatially contiguous information on land surface temperature and surface characteristics throughout the whole urban area. When these data are processed in GIS, they can be compared with land-cover, vegetation, moisture, radiative, and morphology-related variables at the chosen spatial unit.

Recent studies have also used machine learning methods to examine non-linear relationships between urban variables and surface thermal patterns. In this context, Random Forest regression is useful because it can model SUHII and estimate the relative importance of different predictors without assuming a simple linear relation between them. Therefore, the review supports the use of a combined remote sensing, GIS, and machine-learning workflow for comparing Wrocław and Turin[2].

1.2 Introduction to Surface Urban Heat Islands

Early studies of urban climate, starting with the work of Luke Howard, showed that cities can have different thermal conditions from their surrounding rural areas [3]. Later, Oke explained the Urban Heat Island (UHI) as a consequence of changes in the urban surface energy balance [4]. In many traditional UHI studies, the focus is on air temperature within the Urban Canopy Layer, where buildings, streets,

---

[1] Azad Rasul et al., "A Review on Remote Sensing of Urban Heat and Cool Islands," *Land* 6, no. 2 (2017): 38, https://doi.org/10.3390/land6020038; Decheng Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives," *Remote Sensing* 11, no. 1 (2018): 48, https://doi.org/10.3390/rs11010048.

[2] Manar El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania," *Urban Climate* 67 (June 2026): 102872, https://doi.org/10.1016/j.uclim.2026.102872; Guglielmina Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics: A Comprehensive Modelling Approach," *Atmosphere* 15, no. 12 (2024): 1435, https://doi.org/10.3390/atmos15121435.

[3] Luke Howard, *THE CLIMATE OF LONDON*, n.d.

[4] T. R. Oke, "The Energetic Basis of the Urban Heat Island," *Quarterly Journal of the Royal Meteorological Society* 108, no. 455 (1982): 1–24, https://doi.org/10.1002/qj.49710845502.

and vegetation directly affect the local atmosphere [5]. This atmospheric heat island is often stronger at night, when dense urban materials release part of the heat absorbed during the day [6].

With the development of satellite thermal remote sensing, more attention has been given to the Surface Urban Heat Island (SUHI), which is usually assessed through Land Surface Temperature (LST) [7]. LST does not measure air temperature. It represents the radiometric "skin" temperature of surfaces such as asphalt, concrete roofs, bare soil, water, or tree canopies. For this reason, SUHI analysis is especially useful for studying how different surface types contribute to urban heating.

The SUHI effect is mainly linked to the replacement of natural and evaporative surfaces with artificial and impervious materials [8]. Vegetated surfaces can reduce heating through shading and evapotranspiration, while many urban materials absorb and store solar radiation as sensible heat [9]. As a result, exposed artificial surfaces may become much warmer than nearby vegetated or moist surfaces, especially under clear-sky daytime conditions.

This also explains why SUHI and atmospheric UHI do not have exactly the same temporal behavior. SUHI is usually more evident during the daytime and in the warm season, when incoming solar radiation directly heats exposed surfaces [10]. At night, surface temperature patterns are more related to the release of heat stored during the day, while air-temperature UHI effects may remain strong because of reduced cooling in dense urban areas [11].

For this reason, LST should not be interpreted as a direct substitute for air temperature. The two variables describe different parts of the urban thermal environment and are measured in different ways [12]. However, LST remains useful in SUHI studies because it allows the spatial distribution of hot and cool surfaces to be mapped across the whole city. In this thesis, this is important for identifying recurrent

[5] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

[6] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania."

[7] José Antonio Sobrino and Itziar Irakulis, "A Methodology for Comparing the Surface Urban Heat Island in Selected Urban Agglomerations Around the World from Sentinel-3 SLSTR Data," *Remote Sensing* 12, no. 12 (2020): 2052, https://doi.org/10.3390/rs12122052; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[8] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[9] Shushi Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities," *Environmental Science & Technology* 46, no. 2 (2012): 696–703, https://doi.org/10.1021/es2030438; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[10] Marco Morabito et al., "Surface Urban Heat Islands in Italian Metropolitan Cities: Tree Cover and Impervious Surface Influences," *Science of The Total Environment* 751 (January 2021): 142334, https://doi.org/10.1016/j.scitotenv.2020.142334; Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities"; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[11] Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[12] Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

surface hotspots, comparing Wrocław and Turin, and linking thermal patterns with land cover, vegetation, moisture, radiative properties, and urban morphology [13].

1.3 Urban Canopy, Urban Canyon, and Urban Boundary Layer

The previous section introduced SUHI mainly through surface temperature and land-cover change. However, urban heat patterns are also affected by the vertical structure of the city. Buildings, street width, tree height, and the openness of the sky modify how solar radiation reaches the ground, how much heat is stored, and how easily heat can be released back to the atmosphere. For this reason, SUHI analysis should also consider the three-dimensional form of the urban fabric, not only the horizontal distribution of land cover [14].

Following Oke's distinction, urbanization affects two main atmospheric layers: the Urban Canopy Layer (UCL) and the Urban Boundary Layer (UBL) [15]. The Urban Canopy Layer is the lower part of the urban atmosphere, extending from the ground to approximately the average height of buildings and trees. This is the layer where streets, courtyards, façades, roofs, and vegetation directly interact with air flow and radiation exchange.

Within the UCL, the urban canyon is one of the most relevant spatial forms. It refers to the street space enclosed by buildings on one or both sides. The geometry of the canyon affects the local thermal environment because it changes shading conditions, wind movement, and the loss of longwave radiation after sunset [16]. For example, a low Sky View Factor (SVF) indicates that a large part of the sky is obstructed by buildings or trees. This can reduce direct solar exposure during the day, but it can also limit nighttime radiative cooling because less sky is available for heat release [17].

---

[13] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[14] Duo Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity: A Case Study of Xi'an," *Building and Environment* 248 (January 2024): 111072, https://doi.org/10.1016/j.buildenv.2023.111072.

[15] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Oke, "The Energetic Basis of the Urban Heat Island."

[16] David Hidalgo García and Julián Arco Díaz, "Modeling of the Urban Heat Island on Local Climatic Zones of a City Using Sentinel 3 Images: Urban Determining Factors," *Urban Climate* 37 (May 2021): 100840, https://doi.org/10.1016/j.uclim.2021.100840; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity."

[17] Hidalgo García and Arco Díaz, "Modeling of the Urban Heat Island on Local Climatic Zones of a City Using Sentinel 3 Images"; Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity."

The Urban Boundary Layer is located above the canopy and is influenced by the overall roughness, heat storage, and surface characteristics of the urban area [18]. While Canopy Urban Heat Island studies usually rely on air-temperature measurements within the UCL, SUHI studies are based on satellite observations of surface temperature. Therefore, the focus is not the air between buildings, but the radiometric temperature of roofs, roads, vegetation, bare soil, and other exposed surfaces [19].

For this thesis, these concepts are important because surface temperature patterns are not explained only by vegetation or impervious cover. Dense city centers, industrial zones, open residential areas, and green corridors can respond differently to the same weather conditions because their urban form affects shading, heat storage, and surface exposure. This is why morphology-related variables such as SVF, Building Coverage Ratio (BCR), and Surface-to-Volume (S/V) Ratio were included together with spectral and radiative indicators in the SUHII analysis [20].

1.4 Consequences of SUHI

The Surface Urban Heat Island (SUHI) effect has several consequences for urban areas. One of the most direct effects is the increase in cooling demand during warm periods. When urban surfaces store and release more heat, indoor spaces can also become warmer, especially in dense districts with limited vegetation. This increases the use of air conditioning and places additional pressure on energy systems. In turn, cooling systems may contribute to anthropogenic heat release, which can further affect the local thermal environment [21].

SUHI is also relevant for public health and outdoor livability. High surface temperatures do not directly represent air temperature, but they indicate areas where people may experience more difficult outdoor thermal conditions, especially during heatwaves or clear summer days. Prolonged exposure to hot urban environments can increase thermal discomfort and contribute to heat-related health risks, particularly for vulnerable groups such as elderly people, children, and residents with limited access to cooling [22]. As urban populations continue to grow, reducing heat exposure becomes an important part of climate adaptation planning [23].

---

[18] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Oke, "The Energetic Basis of the Urban Heat Island."

[19] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

[20] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity."

[21] Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities"; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[22] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."

[23] Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities."

Another consequence is related to the urban water cycle. The expansion of impervious surfaces reduces infiltration and increases surface runoff. When natural soil and vegetation are replaced by asphalt, concrete, or other sealed materials, less water remains available for evapotranspiration. This weakens one of the main natural cooling processes in cities and contributes to higher land surface temperatures [24].

For these reasons, SUHI analysis is useful for identifying areas where mitigation should be prioritized. GIS, remote sensing, and machine-learning methods can help map surface thermal anomalies and relate them to land cover, morphology, vegetation, moisture, and surface materials [25]. This type of spatial evidence can support planning measures such as increasing vegetation, improving permeability, protecting blue-green infrastructure, and using more reflective surface materials where overheating is recurrent [26].

1.5 Evaluate and mitigate surface urban heat islands

Assessing the Surface Urban Heat Island (SUHI) requires a clear definition of the thermal anomaly being measured. In satellite-based studies, SUHI is commonly derived from Land Surface Temperature (LST), by comparing urban surface temperatures with those of surrounding non-urban or reference areas. The choice of method depends on the data available, the spatial scale of the analysis and the purpose of the study. For example, some studies are based on the maximum thermal contrast, while others use average values to describe the general intensity of the phenomenon [27].

SUHI is usually conveyed by two major indicators:

$$SUHI_{MAX} = LST_{URB-MAX} - LST_{SUR}$$

$$SUHI_{MEAN} = LST_{URB-MEAN} - LST_{SUR}$$

where $LST_{URB-MAX}$ and $LST_{URB-MEAN}$ are the maximum and mean LST of the urban area, respectively, and LST_{SUR} is the mean LST of the surrounding reference area [28].

After the thermal anomaly has been calculated, the next step is to examine which urban and environmental factors are associated with higher surface temperatures. Earlier studies often used simple

[24] Oke, “The Energetic Basis of the Urban Heat Island”; Zhou et al., “Satellite Remote Sensing of Surface Urban Heat Islands.”
[25] Mutani et al., “Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics.”
[26] El Kihel et al., “Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania”; Antonio Esposito et al., “Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves: A Study of Milan and Lecce (Italy),” *Remote Sensing* 16, no. 23 (2024): 4496, https://doi.org/10.3390/rs16234496.
[27] Sobrino and Irakulis, “A Methodology for Comparing the Surface Urban Heat Island in Selected Urban Agglomerations Around the World from Sentinel-3 SLSTR Data.”
[28] Sobrino and Irakulis, “A Methodology for Comparing the Surface Urban Heat Island in Selected Urban Agglomerations Around the World from Sentinel-3 SLSTR Data.”

statistical correlations, for example by comparing LST with the percentage of impervious surface area (ISA)[29]. More recent studies increasingly combine GIS, remote sensing, and machine-learning methods to account for the fact that SUHI is influenced by several variables at the same time. Random Forest regression is useful in this context because it can model non-linear relationships and estimate the relative importance of predictors such as vegetation, moisture, built-up intensity, radiative properties, and urban morphology[30].

Once thermal hotspots and their main drivers are identified, mitigation measures can be interpreted in relation to the physical causes of overheating. Since SUHI is mainly connected to changes in the surface energy balance, planning actions usually focus on reducing heat absorption, increasing evapotranspiration, improving permeability, and modifying urban form where possible[31]. The main groups of interventions discussed in the literature are summarized below.

- Radiative properties: Dark and low-albedo urban materials absorb more solar radiation and can increase LST. Mitigation can include the use of cool materials, high-albedo roofs, and reflective pavements to reduce solar heat absorption on exposed surfaces[32].
- Evaporative and biophysical properties: Vegetation and water help reduce surface heating through shading, evapotranspiration, and latent heat exchange. For this reason, street trees, parks, green roofs, and blue-green infrastructure are often proposed as mitigation measures in overheated urban areas[33].
- Land cover and built-up intensity: Impervious and sealed surfaces are frequently associated with higher LST. Reducing unnecessary surface sealing, increasing permeable areas, and preserving existing green spaces can help limit further thermal stress during urban development or regeneration[34].
- Three-dimensional urban form: Urban geometry affects shading, ventilation, heat storage, and longwave radiation loss. Variables such as Sky View Factor (SVF), building density, and

[29] Huidong Li et al., "A New Method to Quantify Surface Urban Heat Island Intensity," *Science of The Total Environment* 624 (May 2018): 262–72, https://doi.org/10.1016/j.scitotenv.2017.11.360.
[30] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[31] Oke, "The Energetic Basis of the Urban Heat Island."
[32] Esposito et al., "Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[33] Morabito et al., "Surface Urban Heat Islands in Italian Metropolitan Cities"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[34] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Morabito et al., "Surface Urban Heat Islands in Italian Metropolitan Cities."

canyon geometry are therefore relevant for understanding how heat is retained or released in dense urban fabric [35].

In this thesis, these mitigation principles are not treated as simulated design scenarios. Instead, they provide the planning background for interpreting the spatial results obtained from LST, SUHI, SUHII, UTFVI, hotspot recurrence, and Random Forest feature importance. This makes it possible to relate observed thermal patterns to realistic planning options without claiming a before-and-after assessment of regeneration effects.

1.5.1 Key environmental and morphological drivers of SUHI

To model Surface Urban Heat Island Intensity (SUHII), it is necessary to identify variables that describe the main physical conditions influencing land surface temperature. Previous studies show that SUHI patterns are usually related to land cover, vegetation, surface moisture, radiative behavior, and urban morphology [36]. In this thesis, these factors are grouped into three main categories: land-cover and biophysical indicators, built-up and radiative surface properties, and urban morphology variables.

1. **Land cover, vegetation, and moisture indices**

Vegetation and water affect the urban surface energy balance because they modify shading, evapotranspiration, and surface moisture conditions [37]. In satellite remote sensing, these characteristics can be described through spectral indices derived from Landsat bands [38].

- Normalized Difference Vegetation Index (NDVI): NDVI is commonly used to describe vegetation presence, density, and condition by comparing near-infrared and red reflectance [39]. Higher NDVI values are usually associated with cooler surfaces because vegetation provides shade and supports evapotranspiration [40].

---

[35] Esposito et al., "Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[36] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[37] Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities"; Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."
[38] N. Chanpichaigosol et al., "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index," *Global Journal of Environmental Science and Management* (Tehran, Iran) 11, no. 1 (2025): 57–76, https://doi.org/10.22034/gjesm.2025.01.04; Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities."
[39] Chanpichaigosol et al., "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index"; El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania."
[40] Chanpichaigosol et al., "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index"; Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities."

- Normalized Difference Water Index (NDWI): NDWI is used to identify open water bodies and wet surfaces [41]. Water can moderate surface temperature because of its thermal capacity and evaporation, although the cooling effect depends on the size, location, and surrounding urban form [42].
- Proportion Vegetation Index (PVI) and Normalized Difference Moisture Index (NDMI): PVI describes the proportion of vegetation cover, while NDMI provides information on surface moisture and vegetation water content. These indicators are useful for distinguishing areas where vegetation and moisture may contribute to lower surface temperatures [43].

2. Built-up and radiative surface properties

Urban construction materials influence how much solar radiation is absorbed, stored, and emitted by the surface [44]. For this reason, built-up and radiative indicators are commonly used in SUHI studies.

- Normalized Difference Built-up Index (NDBI): NDBI is used to identify built-up and impervious surfaces by comparing shortwave infrared and near-infrared reflectance [45]. Higher NDBI values generally indicate a stronger presence of artificial materials, which are often linked to higher land surface temperatures and stronger SUHI conditions [46].
- Albedo: Albedo represents the fraction of incoming solar radiation reflected by a surface [47]. Dark urban materials usually have lower albedo and absorb more solar energy, while lighter or more reflective surfaces can reduce solar heat absorption [48].

---

[41] Chanpichaigosol et al., "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index"; El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania."

[42] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

[43] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

[44] Peng et al., "Surface Urban Heat Island Across 419 Global Big Cities."

[45] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania."

[46] Chanpichaigosol et al., "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index"; Yantao Xi et al., "Seasonal Surface Urban Heat Island Analysis Based on Local Climate Zones," *Ecological Indicators* 159 (February 2024): 111669, https://doi.org/10.1016/j.ecolind.2024.111669.

[47] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania."

[48] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Cong Wen et al., "Nonlinear Drivers of Canopy and Surface Urban Heat Islands across China's Climate Zones: Marginal and Interaction Effects from Explainable Machine Learning," *Sustainable Cities and Society* 142 (May 2026): 107307, https://doi.org/10.1016/j.scs.2026.107307.

- Emissivity: Emissivity describes the ability of a surface to emit thermal radiation [49]. It is related to the physical properties of surface materials and is therefore relevant for interpreting LST and SUHII patterns, especially in dense or highly sealed urban areas [50].

3. Urban morphology and human-related factors

Beyond two-dimensional land cover, the three-dimensional structure of the city also affects surface heating and cooling. Building density, street geometry, exposed surface area, and sky obstruction can influence shading, ventilation, heat storage, and longwave radiation loss [51].

- Sky View Factor (SVF): SVF expresses the proportion of visible sky from a given location, considering obstruction by buildings, trees, and other elements [52]. Low SVF values can increase shading during the day, but they can also limit radiative heat loss after sunset, especially in narrow urban canyons [53].
- Surface-to-Volume (S/V) ratio and Building Coverage Ratio (BCR): BCR is the ratio of land covered by buildings; S/V Ratio provides information on the relationship between exposed surface area and volume of the buildings. These are variables that help to represent urban compactness and morphology. Higher building coverage can be associated with more sealed surfaces, less open space and changed airflow, but the impact on SUHII depends on the local urban context [54].
- Anthropogenic heat proxies: Population density, nighttime lights, traffic, and energy use are often discussed in the literature as possible proxies for human activity and anthropogenic heat emissions [55]. However, these variables were not included in the final models of this thesis because comparable population-density or human-activity data were not available for both Wrocław and Turin at the required spatial scale. Therefore, the final predictor set focuses on variables that could be produced consistently for both case studies: Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio.

---

[49] Mutani et al., “Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics.”
[50] Mutani et al., “Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics.”
[51] Esposito et al., “Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves”; Wen et al., “Nonlinear Drivers of Canopy and Surface Urban Heat Islands across China’s Climate Zones.”
[52] Hidalgo García and Arco Díaz, “Modeling of the Urban Heat Island on Local Climatic Zones of a City Using Sentinel 3 Images”; Xu et al., “Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity.”
[53] Esposito et al., “Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves”; Hidalgo García and Arco Díaz, “Modeling of the Urban Heat Island on Local Climatic Zones of a City Using Sentinel 3 Images.”
[54] Esposito et al., “Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves”; Xu et al., “Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity.”
[55] Peng et al., “Surface Urban Heat Island Across 419 Global Big Cities”; Wen et al., “Nonlinear Drivers of Canopy and Surface Urban Heat Islands across China’s Climate Zones.”

1.6 Environmental protocols and planning strategies for reducing the SUHI effect

Environmental protocols and planning guidelines can help translate SUHI analysis into practical urban measures. In urban regeneration projects, systems such as LEED and ITACA are often used to evaluate environmental performance, energy use, stormwater management, material choices, and outdoor comfort conditions [56]. In relation to SUHI, these protocols are relevant because they encourage interventions that reduce heat absorption, increase vegetation, improve permeability, and limit unnecessary energy demand.

For this thesis, these protocols are used as planning background rather than as an assessment framework applied directly to the case studies. The analysis does not calculate a LEED or ITACA score. Instead, the role of these protocols is to clarify which types of urban measures are commonly associated with heat mitigation and how they can be interpreted in relation to the spatial results obtained from Landsat, GIS, and Random Forest modeling.

The Local Climate Zone (LCZ) classification is also relevant for planning interpretation. LCZ groups urban and natural areas according to their built form, land cover, surface materials, and thermal behavior [57]. This makes it useful for reading thermal patterns in relation to urban morphology. In this thesis, LCZ is not used as a predictor in the Random Forest models. It is used only as a rule-based interpretive layer to support the discussion of planning priorities in different urban contexts.

The main planning themes related to SUHI mitigation can be summarized as follows:

- Radiative properties and materials: Urban surfaces with low albedo absorb more solar radiation and can contribute to higher LST. Environmental protocols therefore often support the use of cool materials, reflective roofs, and lighter pavements, especially in dense or sealed areas where increasing vegetation is difficult [58].
- Urban morphology: The morphology of the built environment affects shading, ventilation and longwave radiation exchange. Parameters such as, Sky View Factor (SVF), Building Coverage Ratio (BCR) and the shape of the urban canyon can account for the difference of heat retention within dense [59]. Planning interventions in these areas include improving ventilation corridors, reducing excessive surface sealing and avoiding forms which increase heat accumulation.

---

[56] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[57] I. D. Stewart and T. R. Oke, "Local Climate Zones for Urban Temperature Studies," Bulletin of the American Meteorological Society, *Bulletin of the American Meteorological Society* 93, no. 12 (2012): 1879–900, https://doi.org/10.1175/BAMS-D-11-00019.1; Xi et al., "Seasonal Surface Urban Heat Island Analysis Based on Local Climate Zones."
[58] Antonio Esposito et al., "A Multi-City Statistical Modelling of Surface Urban Heat Island: Application to Italian Cities," *Urban Climate* 64 (December 2025): 102717, https://doi.org/10.1016/j.uclim.2025.102717; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[59] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity."

- Permeability and green infrastructure: One of the most common ways to reduce surface overheating is to increase the amount of permeable surfaces and vegetation. Street trees, green roofs, vertical greenery, parks and planted courtyards can cool surface temperatures through shading and evapotranspiration.[60]. These measures are especially relevant where recurrent hotspots meet sealed or poorly vegetated urban fabric.
- Blue infrastructure: Rivers, canals, ponds, and other water surfaces can influence local thermal conditions through evaporation and heat storage. Their cooling effect depends on their size, continuity, exposure, and connection with surrounding green areas [61]. For this reason, blue infrastructure is most useful when it is planned together with green corridors and permeable open spaces.
- Anthropogenic heat management: Energy use, traffic, and industrial activities can add heat to the urban environment. Although human activity variables were not included in the final Random Forest models of this thesis, the planning literature still recognizes the importance of reducing waste heat through energy-efficient buildings, public transport, bioclimatic design, and lower-emission urban systems [62].

Overall, this section provides the planning context for the later interpretation of SUHI and SUHII results. The thesis does not test the performance of individual mitigation scenarios. Rather, it uses the mapped thermal patterns, hotspot recurrence, feature-importance results, and LCZ-based interpretation to discuss which types of planning measures are more suitable for Wrocław and Turin.

2. Studying surface urban heat islands: the case of Turin and Wrocław

This chapter introduces the two case-study cities used in the thesis: Wrocław, Poland, and Turin, Italy. The comparison is based on the idea that SUHI patterns are influenced not only by land cover and surface materials, but also by local climate, topography, urban form, and planning history. Wrocław and Turin provide two different urban and environmental settings for testing the same analytical workflow.

Wrocław is located in the Silesian Lowlands and within the Oder River valley. Its climate has temperate oceanic characteristics with continental influences, and its urban thermal patterns are affected by the relationship between built-up areas, river corridors, vegetation, and peri-urban land. Turin is located in the Po Valley, at the foot of the Alps, within a more enclosed basin setting. This geographical condition

[60] Morabito et al., "Surface Urban Heat Islands in Italian Metropolitan Cities"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[61] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity."
[62] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

can favor heat accumulation and reduced air circulation, especially in stable summer weather conditions.
Thus, the two cities are not compared because they are similar, but because they offer the possibility to test the same SUHI methodology under different urban and climatic conditions. In both cases, the analysis examines the role of land cover, vegetation, moisture, surface radiative properties and urban morphology in shaping surface temperature patterns.

The workflow was first developed and checked for Wrocław and then applied to Turin using the same main processing steps. This sequence was used to keep the comparison consistent and to reduce the possibility that differences between the cities were caused by changes in the method. The chapter therefore connects the literature review with the following methodological chapters, where the pre-modeling variables, multi-seasonal spatial assessment, and warm-season SUHII modeling are described in detail.

2.1 Main purposes of the SUHI analysis in Turin and Wrocław

The main purpose of this thesis is to analyze how surface thermal patterns vary between Wrocław and Turin and to identify which spatial variables help explain Surface Urban Heat Island Intensity (SUHII) in each city. The work does not only map high and low temperature areas. It also relates these patterns to land cover, vegetation, moisture, radiative properties, and urban morphology.

The comparison between Wrocław and Turin is useful because the two cities have different climatic and geographical settings. Wrocław is located in the Silesian Lowlands and the Oder River valley, with temperate oceanic conditions and continental influences. Turin is located in the Po Valley, close to the Alpine foothills, where the basin-like setting can favor heat accumulation during stable weather. These differences make it possible to test whether the same methodological workflow can describe SUHI patterns in two different urban contexts.

The analysis has four specific aims. The first is to produce a multi-seasonal spatial assessment of LST, SUHI, SUHII, UTFVI, and hotspot recurrence for both cities. The second is to prepare comparable census-section datasets from Landsat-derived indicators, morphology variables, and meteorological context data. The third is to model warm-season SUHII using Random Forest regression and to compare the relative importance of the predictors in Wrocław and Turin. The fourth is to interpret the spatial and modeling results in relation to planning priorities, including a rule-based LCZ reading used only as a planning interpretation layer.

The workflow was developed first for Wrocław and then applied to Turin using the same main processing steps. This sequential approach was used to improve comparability between the two case studies. It also reduces the risk that observed differences in SUHII patterns or predictor importance are caused by changes in data processing rather than by local urban and environmental conditions.

Human activity and population density are recognized in literature as possible contributors to urban heat. However, they were not included as predictors in this thesis because comparable data were not available for both cities at the required spatial scale. For this reason, the final modeling focuses on variables that could be produced consistently for Wrocław and Turin: Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio.

2.2 Visualizing Thesis Methodology: Flowchart Overview

This section summarizes the methodological sequence followed in the thesis. The workflow starts with the literature review and framing of the case-study, which define the main concepts, variables and planning questions related to SUHI. Then the data preparation follows including the selection of Landsat 8/9 scenes, the calculation of spectral and radiative indicators in Google Earth Engine, the aggregation of the raster values to census sections in QGIS and the processing of ERA5-Land meteorological data in Python. The analysis is divided into three major levels following the pre-modeling. The first level is a multiseasonal spatial assessment where surface thermal patterns are compared between seasons using LST, SUHI, SUHII, UTFVI, and hotspot recurrence. The second level is the Random Forest modeling of SUHII for the warm season based on the final predictor set of Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR and S/V Ratio. The third level is the planning interpretation, where the discussion of the spatial and modeling results is combined with a rule-based LCZ reading.

The same general workflow was first developed for Wrocław and then applied to Turin. This sequence was used to keep the processing steps comparable between the two case studies. Figure 2.1 presents the complete structure of the methodology, from data preparation to spatial assessment, predictive modeling, and planning interpretation.

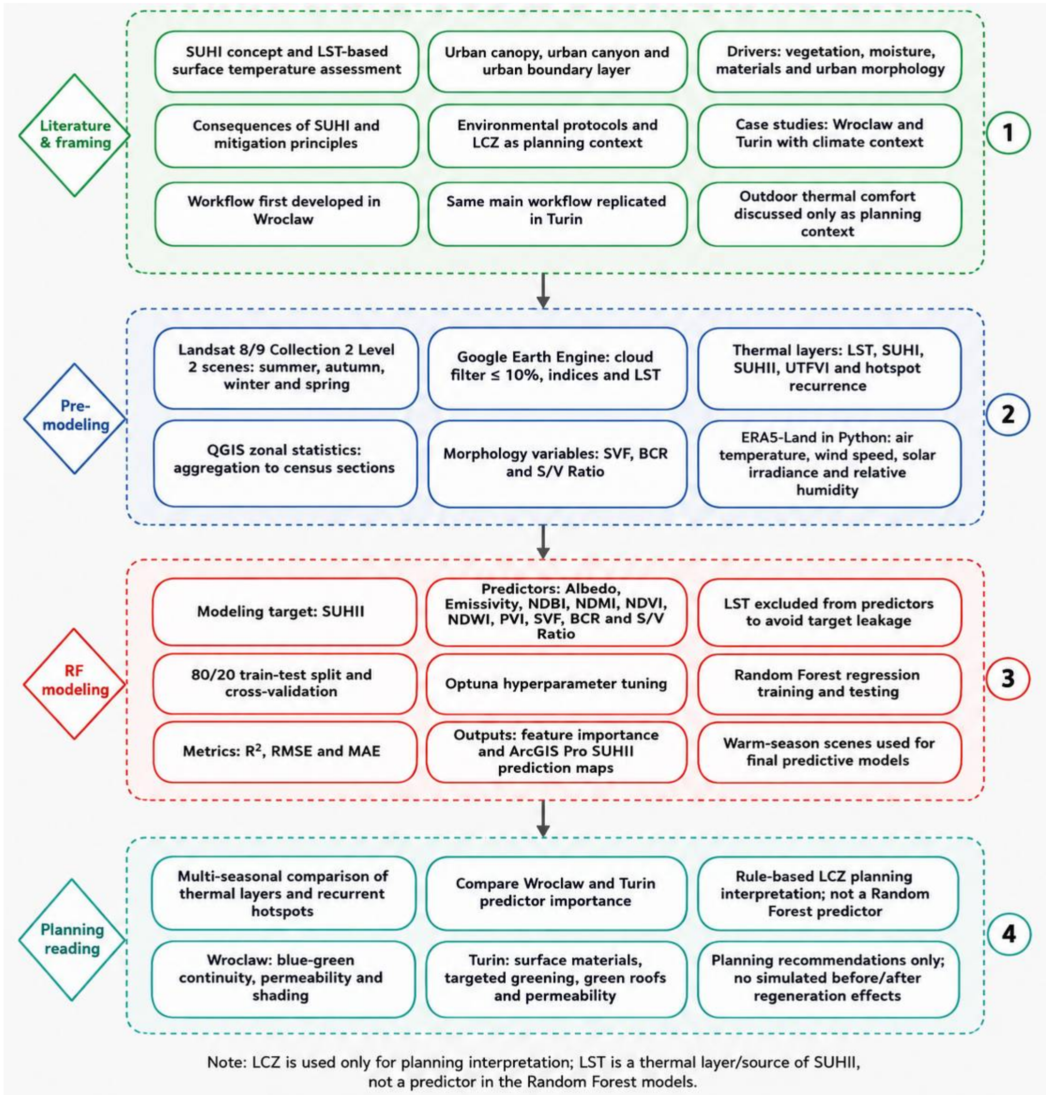


Figure 2.1. General workflow of the integrated modeling approach: from data acquisition to warm-season predictive modeling and planning interpretation.

2.3 Understanding Climate Context: A Brief Overview

Urban surface temperature patterns are not determined only by land cover, building density, or surface materials. They are also affected by the climatic setting in which the city is located. Seasonal temperature variation, humidity, solar radiation, wind conditions, and topography can influence how strongly surfaces heat up and how quickly they cool. For this reason, a short climate comparison is needed before interpreting the SUHI results for Wrocław and Turin.

Wrocław is located in the Silesian Lowlands of Central Europe and within the Oder River valley. Its climate can be described as temperate oceanic with continental influences. This setting produces marked seasonal differences, with cold winters and warm summers, and it provides a useful context for examining how surface thermal patterns change across the selected Landsat scenes [63].

Turin is located in northwestern Italy, at the foot of the Alps and within the Po Valley. Its humid subtropical climate (Cfa), together with the basin-like geographical setting, can favor heat accumulation and reduced air circulation during stable weather conditions, especially in summer [64]. This makes the city relevant for comparing how surface overheating behaves in a different climatic and topographic context from Wrocław.

This section does not aim to provide a detailed climatological analysis of the two cities. Its purpose is to give the basic meteorological and geographical background needed to interpret the multi-seasonal spatial assessment and the warm-season SUHII modeling presented in the following chapters. The main comparative parameters are summarized in Table 2.1.

Table 2.1. Comparative climatic and geographic parameters for the case studies of Wrocław and Turin.

Parameter	Wrocław (Poland)	Turin (Italy)
Geographic context	Silesian Lowlands / Oder River valley	Po Valley / Alpine foothill basin
Köppen classification	Temperate oceanic with continental influences (Cfb / Dfb)	Humid subtropical (Cfa)
Elevation (a.s.l.)	~120 m	~239 m
Mean annual air temperature	~9.7 °C	~12.8 °C
Mean annual precipitation	~590 mm	~980 mm
Warmest month	July	July
Coldest month	January	January
Representative summer temp.	~24–25 °C	~28–31 °C
Representative winter temp.	~ −3 to −1 °C	~ −1 to 1 °C

[63] Melika Tasan et al., “Long-Term Dynamics of Urban Heat Island and Hot Spots in Wrocław: A 25-Year Satellite-Based Analysis Using Machine Learning,” *Sustainable Cities and Society* 132 (September 2025): 106797, https://doi.org/10.1016/j.scs.2025.106797.
[64] Mutani et al., “Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics.”

Key climatic features	Continental influence, river-valley setting, marked seasonality	Basin enclosure, Alpine influence, stagnant conditions

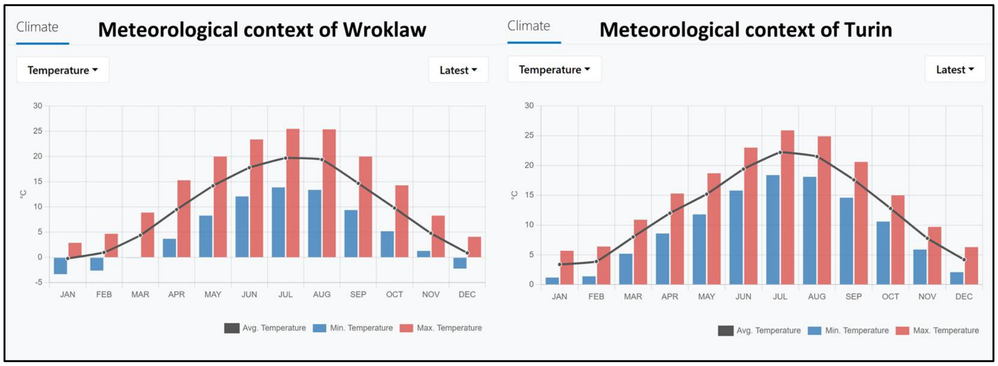


Figure 2.2. Monthly average, minimum, and maximum air temperatures for Wrocław and Turin, used as climatic context for the multi-seasonal SUHI assessment and warm-season SUHII modeling. Data source: Meteostat.

3. Pre-modeling: Assessment of Environmental Variables

3.1 Data collection and processing

The workflow was prepared first in Wroclaw and then repeated in Turin by the same main steps. Four Landsat scenes per city were chosen to represent summer, autumn, winter and spring conditions. The scenes were used for multi-seasonal spatial assessment of LST, SUHI, SUHII, UTFVI and hotspot recurrence. Then, for each city, a Random Forest model of SUHII was constructed using the warm-season representative scene. Table 3.1 lists the selected acquisition dates, seasonal labels, and sensor information. During the Google Earth Engine filtering stage, a cloud-cover threshold of 10% or lower was used to select suitable scenes.

Table 3.1. Overview of selected Landsat scenes for multi-seasonal spatial assessment in Wrocław and Turin.

City	Date	Season label	Sensor
Wrocław	26 July 2024	Summer	Landsat 8 / 9
	20 September 2024	Autumn	
	18 January 2025	Winter	
	5 May 2025	Spring	
Turin	28 June 2024	Summer	Landsat 8 / 9
	10 November 2024	Autumn	
	13 January 2025	Winter	
	30 May 2025	Spring	

Environmental and thermal variables were derived from Landsat 8/9 Collection 2 Level 2 imagery in Google Earth Engine (GEE). For each city and date, the image collection was filtered by Area of Interest (AOI), acquisition period, and cloud-cover metadata. When more than one image was available, the least cloudy scene was selected. This step reduced the influence of cloud contamination and allowed the same processing logic to be applied to both case studies. Figure 3.1 summarizes the GEE workflow used for scene selection, variable calculation, and raster export.

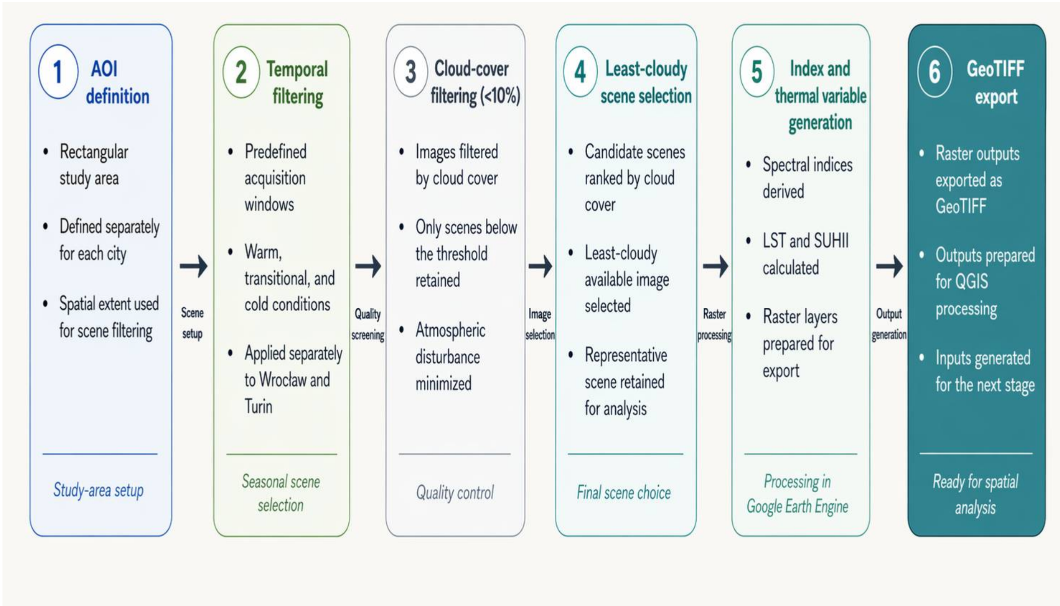


Figure 3.1. Google Earth Engine (GEE) workflow for filtering satellite data, generating variables, and exporting rasters.

The GEE workflow produced the main raster layers used in the thesis. These included Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI), Albedo, Proportion of Vegetation Index (PVI), Emissivity, Land Surface Temperature (LST), and Surface Urban Heat Island Intensity (SUHII). LST was calculated from the thermal band with emissivity correction, while SUHII was calculated as the standardized LST value. In the modeling phase, SUHII was used as the target variable, while LST was not used as a predictor. Table 3.2 reports the full names, source bands, and analytical roles of the Landsat-derived variables.

Table 3.2. Overview of environmental and thermal variables derived from Landsat 8/9 data.

Variable	Full name	Data source/bands	Role in analysis
NDVI	Normalized Difference Vegetation Index	Landsat 8/9 L2; B5/B4	Vegetation density
NDBI	Normalized Difference Built-up Index	Landsat 8/9 L2; B6/B5	Built-up intensity
NDWI	Normalized Difference Water Index	Landsat 8/9 L2; B3/B5	Water-related surface signal
NDMI	Normalized Difference Moisture Index	Landsat 8/9 L2; B5/B6	Moisture condition
Albedo	Surface Albedo	Landsat 8/9 reflectance bands	Surface reflectivity

PVI	Proportion of Vegetation Index	Derived from NDVI	Fractional vegetation cover
Emissivity	Land Surface Emissivity	Derived from PVI	Thermal emissive property
LST	Land Surface Temperature	Thermal band + emissivity correction	Surface temperature
SUHII	Surface Urban Heat Island Intensity	Standardized LST (computed in GEE)	Modelling target

After the raster layers were exported from GEE, they were imported into QGIS for spatial aggregation. Zonal statistics were used to calculate the mean value of each raster variable within each census-section unit. This converted the pixel-based outputs into a polygon-based attribute table that could be used for later data preparation and Random Forest modeling. Figure 3.2 shows the raster-to-table aggregation process.

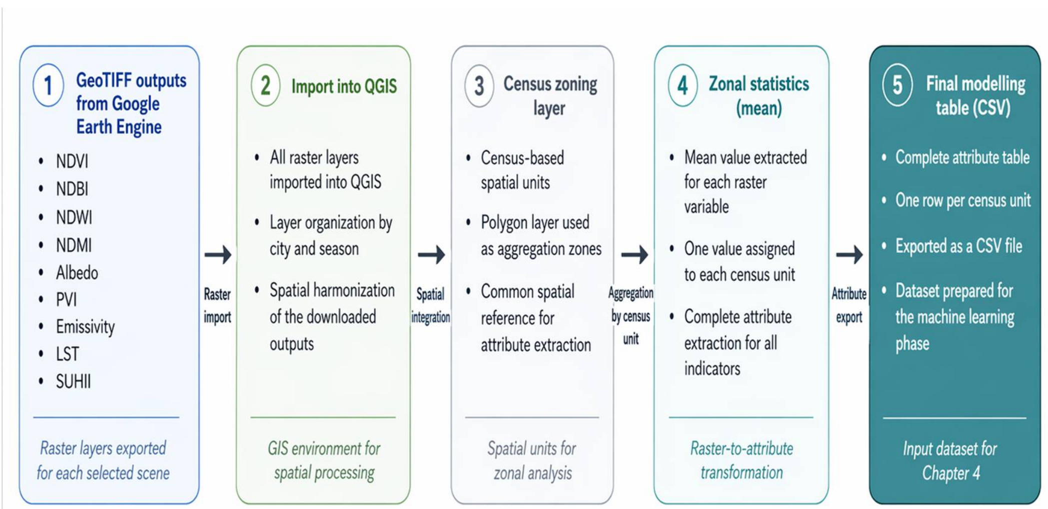


Figure 3.2. Raster-to-table aggregation workflow: from Landsat-derived indices to the census-based structured modeling dataset.

The resulting attributes were exported as a single CSV file and used as the main input for the machine-learning analysis described in Chapter 4 [65]. Additional QGIS processing was also carried out before modeling to support the multi-seasonal spatial assessment. UTFVI was calculated from the LST layer, while SUHI was calculated using the rural-reference LST contrast. The SUHI and UTFVI layers were

[65] Chanpichaigosol et al., "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index."

then used to identify and interpret thermal stress and hotspot patterns across the selected dates [66]. These steps provided the spatial assessment layer of the thesis before the warm-season SUHII modeling.

3.1.1 Google Earth Engine implementation and personal contribution

Purpose and personal contribution.

The first processing step of the thesis was carried out in Google Earth Engine (GEE). In this phase, the Landsat scenes were not only downloaded, but also filtered, checked, processed, and converted into raster layers through a custom GEE script. This script was prepared first for Wrocław and then adapted for Turin, using the same main logic for both case studies.

The purpose of this step was to create the remote-sensing variables needed for the later GIS and modeling phases. These variables included vegetation, moisture, built-up, radiative, and thermal indicators. The GEE workflow therefore formed the starting point of the full analysis, before the raster outputs were transferred to QGIS for zonal statistics and dataset preparation.

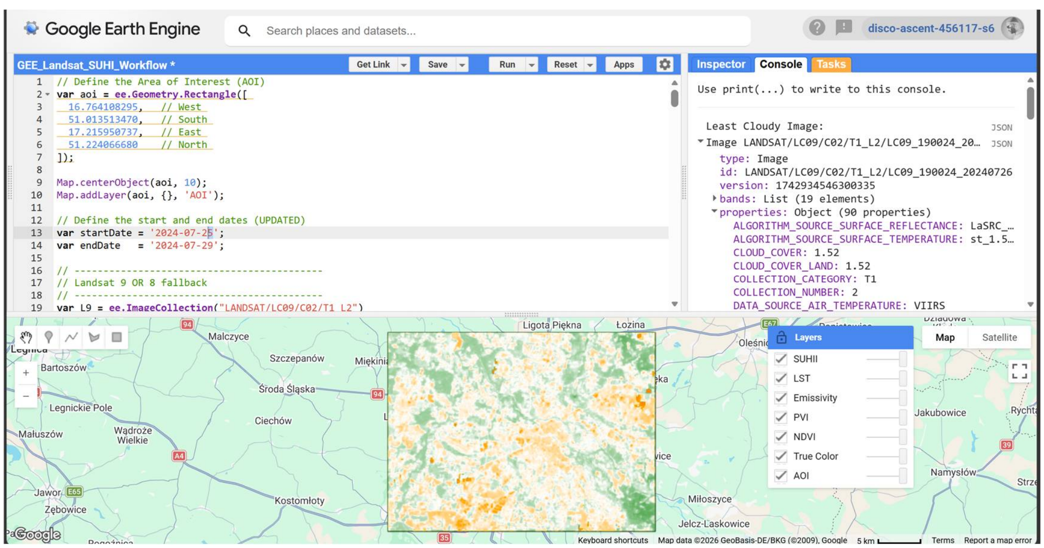


Figure 3.3. Google Earth Engine implementation of the Landsat preprocessing and raster-generation workflow. screenshot shows the custom script, the selected Landsat scene metadata, and the Area of Interest, and the produced raster layers used prior to transfer to the GIS workflow.

AOI, scene selection, date window.

For each selected city and acquisition date, the script defined the Area of Interest (AOI) and date window for searching the Landsat image collection. Landsat 8/9 Collection 2 Level 2 scenes were

---

[66] Yiğitalp Kara et al., "Multi-Index Assessment of Surface Urban Heat Island (SUHI) Dynamics in Samsun Using Google Earth Engine," *Atmosphere* 16, no. 6 (2025): 712, https://doi.org/10.3390/atmos16060712.

selected by spatial extent, date, and cloud-cover metadata. The available scenes were then inspected in the GEE Console.

When more than one suitable scene was available, the image collection was sorted by cloud-cover value and the least-cloudy scene was selected. This step was used to reduce cloud contamination and to keep the input image as clear as possible for the calculation of LST and SUHII.

```
// Define the Area of Interest (AOI)
var aoi = ee.Geometry.Rectangle([
  -9.266315101,   // West
  38.674435684,   // South
  -9.047615918,   // East
  38.813458802    // North
]);

// Add AOI to the map and set the view
Map.centerObject(aoi, 10);
Map.addLayer(aoi, {}, 'AOI');

// Define the start and end dates
var startDate = '2024-07-01';
var endDate = '2024-07-10';

// Apply scaling factor function
function applyScaleFactors(image) {
  var opticalBands =
image.select('SR_B.').multiply(0.0000275).add(-0.2);
  var thermalBands =
image.select('ST_B.*').multiply(0.00341802).add(149.0);
  return image.addBands(opticalBands, null, true)
              .addBands(thermalBands, null, true);
}
```

Figure 3.4. GEE code excerpt for AOI definition, temporal filtering, and least-cloudy Landsat scene selection.

Preprocessing and scaling factors.

The selected Landsat image was pre-processed with the official Collection 2 Level 2 scale factors. The optical surface-reflectance bands were rescaled using the multiplicative factor and additive offset provided for the Landsat product. The thermal band was also converted using the appropriate surface temperature scale factor.

The script included a cloud and shadow mask using QA_PIXEL band. This made it possible to remove cloud affected and shadow contaminated pixels when needed. After these preprocessing steps the image was clipped to the AOI and used to compute the raster variables.

```
// Cloud masking function
function maskClouds(image) {
  var cloudBitMask = (1 << 3);
  var shadowBitMask = (1 << 4);
  var qa = image.select('QA_PIXEL');
  var clMask = qa.bitwiseAnd(cloudBitMask).eq(0);
  var shMask = qa.bitwiseAnd(shadowBitMask).eq(0);
  return image.updateMask(clMask.and(shMask));
}

// Print the names of selected scenes
var imageNames = ee.ImageCollection("LANDSAT/LC09/C02/T1_L2")
                  .filterBounds(aoi)
                  .filterDate(startDate, endDate)
                  .distinct('LANDSAT_SCENE_ID');

print('Selected LANDSAT Scenes:',
imageNames.aggregate_array('LANDS_SCENE_ID'));

// Select the least cloudy image
var leastCloudyImage =
ee.ImageCollection("LANDSAT/LC09/C02/T1_L2")
                        .filterBounds(aoi)
                        .filterDate(startDate, endDate)
                        .sort('CLOUD_COVER')  // Sort by cloud
cover
                        .first();
// Select the first (least cloudy) image

print('Least Cloudy Image:', leastCloudyImage);

// Preprocess the least cloudy image
var leastCloudyVis = applyScaleFactors(leastCloudyImage)  //
Apply scaling factors
                    //
.updateMask(maskClouds(leastCloudyImage))

// Mask clouds and shadows
                     .clip(aoi);
// Clip the image to the AOI
```

Figure 3.5. GEE preprocessing code used to apply Landsat scale factors, prepare cloud and shadow masking, and clip the selected image to the study area.

Raster variables and link to Table 3.2.

After preprocessing, the variables listed in Table 3.2 were calculated from the Landsat bands and intermediate raster layers. These included NDVI, NDBI, NDWI, NDMI, Albedo, PVI, Emissivity, LST, and SUHII. The formulas and analytical roles of the variables are summarized in Table 3.2, while the screenshots in this section show how the calculations were implemented in GEE.

Thermal calculation chain: NDVI, PVI, emissivity, LST, and SUHII.

The thermal part of the workflow followed a defined sequence. First, NDVI was calculated from the red and near-infrared bands. Then, the minimum and maximum NDVI values within the AOI were extracted using regional reducers. These values were used to calculate PVI, which was then used to derive surface emissivity.

The emissivity layer was included in the correction of the Landsat thermal band to obtain LST. After LST was calculated, the AOI-based mean and standard deviation of LST were extracted. These two values were then used to standardize each LST pixel into SUHII:

$$SUHII\_{i} = (LST\_{i} - LST_{mean}) / LST_{std}$$

In this way, the final SUHII raster was derived from the same sequence of intermediate layers for both cities. SUHII was later used as the modeling target, while LST was kept for spatial assessment and was not used as a Random Forest predictor.

```
// Calculate NDVI
var ndvi = leastCloudyVis
  .normalizedDifference(['SR_B5', 'SR_B4'])
  .rename('NDVI');

// Calculate NDVI min and max within AOI
var ndviMin = ee.Number(ndvi.reduceRegion({
  reducer: ee.Reducer.min(),
  geometry: aoi,
  scale: 30,
  maxPixels: 1e9
}).values().get(0));

var ndviMax = ee.Number(ndvi.reduceRegion({
  reducer: ee.Reducer.max(),
  geometry: aoi,
  scale: 30,
  maxPixels: 1e9
}).values().get(0));

// Calculate Proportion Vegetation Index (PVI)
var pvi = ndvi
  .subtract(ndviMin)
  .divide(ndviMax.subtract(ndviMin))
  .pow(ee.Number(2))
  .rename('PVI');

// Calculate surface emissivity
var em = pvi
  .multiply(ee.Number(0.004))
  .add(ee.Number(0.986))
  .rename('EM');

// Select thermal band
var thermal = leastCloudyVis
  .select('ST_B10')
  .rename('thermal');
```

Figure 3.6. GEE code sequence for deriving NDVI, PVI, and surface emissivity from the preprocessed Landsat image.

```
// Calculate Land Surface Temperature (LST)
var lst = thermal.expression(
  '(TB / (1 + (0.00115 * (TB / 1.438)) * log(em))) - 273.15', {
    'TB': thermal.select('thermal'),
    'em': em
  }).rename('LST');

// Calculate LST mean and standard deviation
var lstStats = lst.reduceRegion({
  reducer: ee.Reducer.mean().combine({
    reducer2: ee.Reducer.stdDev(),
    sharedInputs: true
  }),
  geometry: aoi,
  scale: 30,
  maxPixels: 1e9
});

var lstMean = ee.Number(lstStats.get('LST_mean'));
var lstStdDev = ee.Number(lstStats.get('LST_stdDev'));

// Calculate Surface Urban Heat Island Intensity (SUHII)
var suhii = lst.expression(
  '(LST - LSTmean) / LSTstddev', {
    'LST': lst.select('LST'),
    'LSTmean': lstMean,
    'LSTstddev': lstStdDev
  }).rename('SUHII');
```

Figure 3.7. GEE code sequence for calculating LST and standardizing it into SUHII using the AOI-based LST mean and standard deviation.

Export to QGIS.

The last step in GEE was the export of the generated raster layers. The selected outputs were saved as GeoTIFF files to Google Drive at Landsat spatial resolution. They were then imported into QGIS for spatial checking, NoData management, reprojection or harmonization where required, and aggregation to census-section units. These exported rasters provided the remote-sensing input for the QGIS workflow described in the next section.

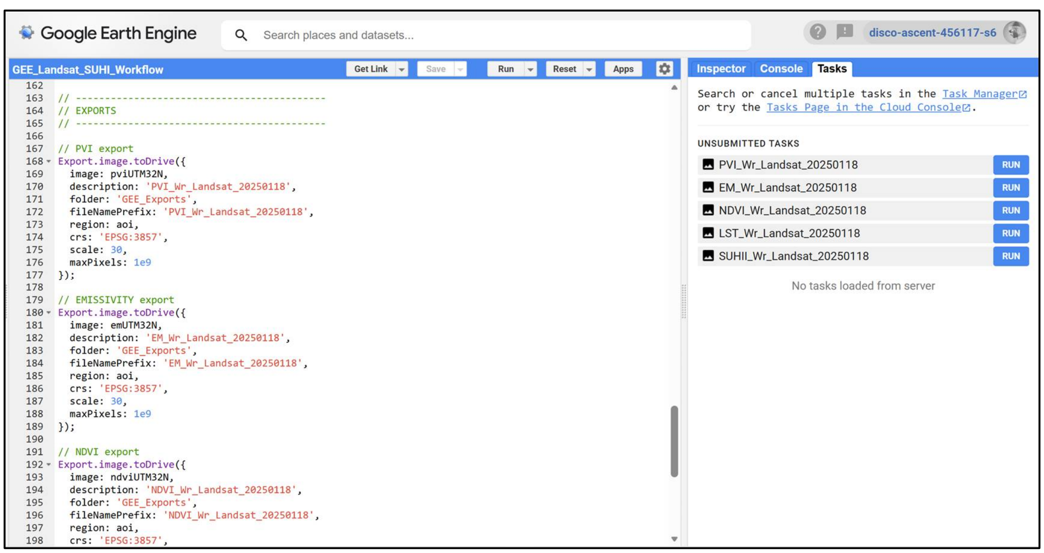


Figure 3.8. GEE export code used to save the generated raster layers as GeoTIFF files for subsequent QGIS processing.

3.1.2 QGIS zonal statistics, morphological variables, and preparation of the final modeling dataset

QGIS zonal aggregation framework.

After the raster layers were generated in Google Earth Engine, they were imported into QGIS for spatial checking and aggregation. The census-section layer for each case-study city was used as the common spatial unit. This choice made it possible to convert the pixel-based Landsat outputs into polygon-based attributes that could later be used in the modeling dataset.

The purpose of this step was to prepare a consistent table in which each census section contained the mean values of the selected raster variables. In this way, the remote-sensing outputs could be linked to the same spatial units used for the morphology variables and the Random Forest analysis.

Figure 3.9. QGIS workspace used to organize the Landsat-derived raster layers and the census-section polygon layer before zonal aggregation.

Sequential zonal statistics on census sections.

The raster to polygons aggregation was performed using the Zonal Statistics tool in QGIS. Each raster derived from Landsat was independently processed with the target layer being the census-section polygons. After each run the calculated mean value was added as a new attribute to the same census-based layer. The same was done for the selected raster variables: vegetation, moisture, built-up, radiative and thermal indicators. This yielded a census-section attribute table with the Landsat-derived variables required for the spatial assessment and subsequent modelling steps.

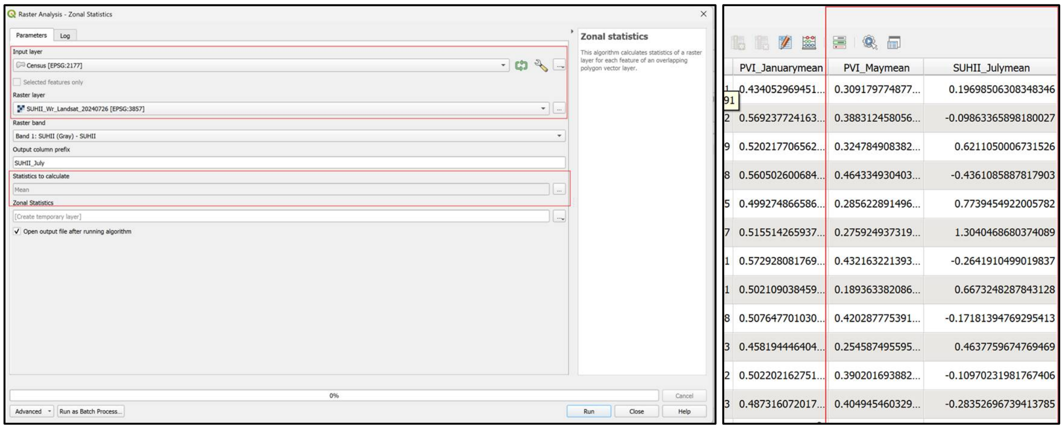


Figure 3.10. QGIS zonal statistics workflow used to aggregate Landsat-derived raster values to census-section units. The figure combines the Zonal Statistics tool interface with an example of the updated attribute table, showing how each raster-derived variable was progressively added to the census-based modeling dataset.

Iterative attribute enrichment across seasonal scenes.

The same zonal statistics procedure was applied to the four selected seasonal Landsat scenes for each city. This allowed the summer, autumn, winter, and spring outputs to be stored and compared within the same census-section structure. The seasonal variables were used mainly to support the multi-seasonal spatial assessment of LST, SUHI, SUHII, UTFVI, and hotspot recurrence.

For the Random Forest phase, the modeling focus was placed on the warm-season representative scene, when the surface thermal contrast was clearer. This distinction is important because the thesis does not treat all seasons as separate predictive models. Instead, the seasonal layers support spatial comparison, while the warm-season dataset provides the main input for SUHII modeling.

Final_DataSet — Features Total: 645, Filtered: 645, Selected: 0

	NDBI_Maymean	NDMI_Julymean	NDMI_Septembermean	NDMI_Januarymean	NDMI_Maymean	NDVI_Julymean	NDVI_Septembermean	NDVI_Januarymean	NDVI_Maymean	ND
1	-0.21836694786453262	0.18231574825220528	0.18949943314703746	-0.04494927194954458	0.21836694786453262	0.5064290921041451	0.47468976661938866	0.28998062798907887	0.5380706911513565	-0.41
2	-0.25423232898836134	0.18479114529442917	0.19408811636645396	0.10046960519699562	0.25423232898836134	0.6377317663456356	0.6496314276067865	0.5035168520659776	0.6122479251317182	-0.60
3	-0.2107134958761826	0.1568232088473469	0.17756915204587304	0.0621260726338026	0.2107134958761826	0.560141806831933	0.5735455964292798	0.43811730979790064	0.559334732573427	-0.55

Figure 3.11. Census-section attribute table progressively enriched through the seasonal zonal statistics workflow. The table shows raster-derived variables from multiple Landsat acquisition dates stored in a single polygon layer, allowing the four seasonal scenes to be compared and prepared for subsequent Python-based modeling.

Calculation of variables related to morphology.

Morphology-related indicators were also calculated in QGIS, based on the building footprint and census-section layers, in addition to the Landsat-derived variables. First, the Building Coverage Ratio (BCR) was calculated, which required the size of each census section. Then building footprint area, perimeter and height were used to calculate building level geometry values.

Building Volume was derived as an intermediate value from footprint area and building height. It was used to support the calculation of the Surface-to-Volume Ratio (S/V Ratio), which describes the relationship between exposed building surface and built volume. The building footprint areas were then summarized by census section and joined to the census-section layer. BCR was calculated as the ratio between the total building footprint area within a census section and the total area of that census section.

Through this process, the census-based layer was enriched with the morphology variables used in the thesis: SVF, BCR, and S/V Ratio. Building Volume was part of the calculation workflow, but it was not used as a final Random Forest predictor.

```
-- BCR
"sum_Area_bld" / $area

-- Volume
$area * "Height"

-- S/V
((2 * $area) + ("Perimeter" * "Height")) / ($area * "Height")
```

Figure 3.12. QGIS field-calculation formulas used for morphology-related variables. Census-section area was used as the reference area for BCR, while building footprint area, perimeter, and height were used in the building layer to calculate Building Volume and Surface-to-Volume Ratio before joining the summarized attributes to the final census-based layer.

Completion of the final census-based attribute table.

After the zonal statistics and morphology calculations, the final census-section layer contained the main variables required for the analysis. The Landsat-derived variables described surface conditions such as vegetation, moisture, built-up intensity, radiative behavior, and surface temperature. The morphology variables described the built-form characteristics of each census section.

The completed table was checked to ensure that the variables were stored with consistent names and that each record corresponded to one census-section unit. This step was necessary before exporting the table for Python processing and Random Forest modeling.

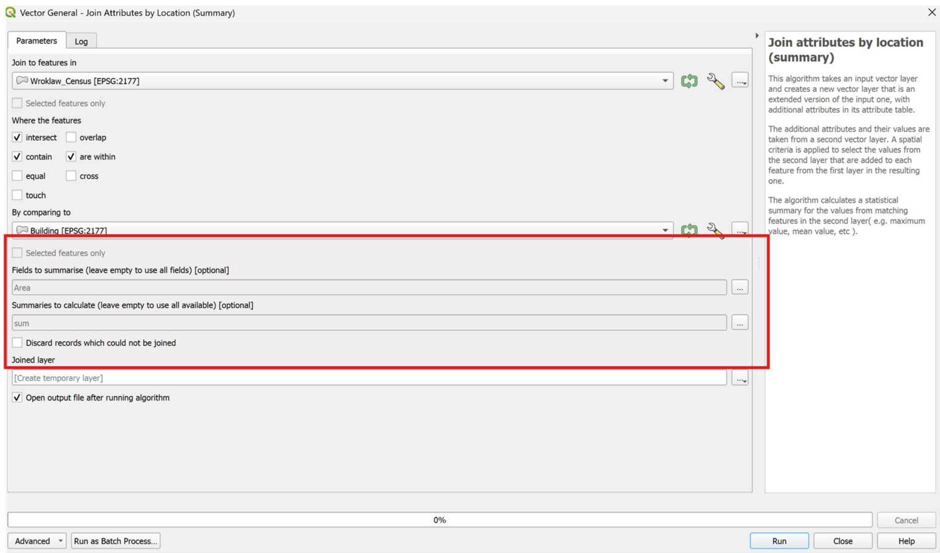


Figure 3.13. QGIS join operation used to complete the final census-based attribute table before export.

Export of the final modeling dataset.

The final census-based attribute table was exported from QGIS as a CSV file. This file was used in the following stages for data cleaning, variable selection, train-test splitting, model tuning, and Random Forest regression. In the modeling phase, SUHII was used as the target variable. LST was retained for spatial assessment and interpretation, but it was not used as a predictor in order to avoid target leakage.

The final predictor set used consistently for both cities was: Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio.

3.1.3 Python processing of ERA5-Land meteorological data

ERA5-Land data source and meteorological input preparation.

Meteorological data were incorporated to characterize the atmospheric conditions on the selected Landsat acquisition dates. The data were downloaded from the Copernicus Climate Data Store ERA5-Land hourly reanalysis. One NetCDF file was downloaded for each city and for each Landsat scene that matched the date and spatial extent.

ERA5-Land was chosen because it provides gridded data in the same structure for Wrocław and Turin. This helped to keep the meteorological input comparable between the two case studies. The downloaded files contained the variables required for the thesis: 2 m air temperature, 2 m dewpoint temperature, 10 m wind components and surface solar radiation downwards.

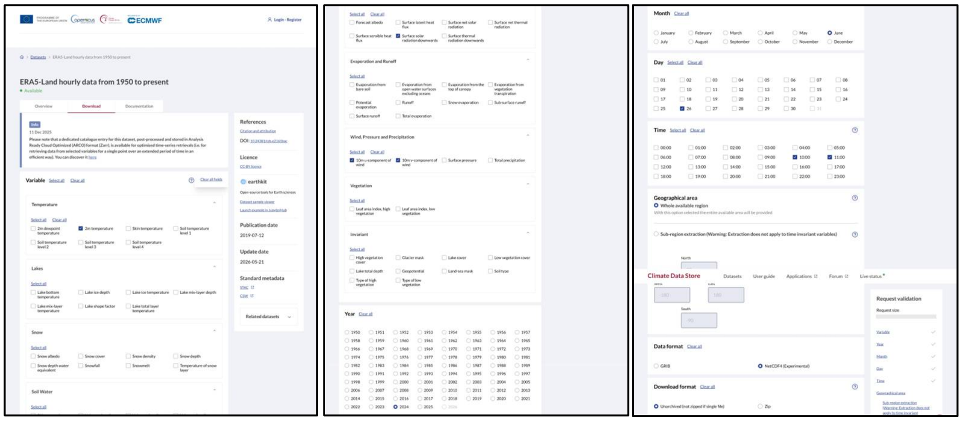


Figure 3.14. Copernicus Climate Data Store interface used to retrieve ERA5 - Land hourly NetCDF data for the selected Landsat acquisition dates and case-study extents.

Source: Copernicus Climate Data Store, ERA5-Land hourly data from 1950 to present.

Python file organization and NetCDF inspection.

The NetCDF files were processed in Python. Each file was linked to its city, acquisition date, and season before the calculations were carried out. The script then opened each file, checked the available variables and dimensions, and verified that the required ERA5-Land fields were present.

This file structure reduced the risk of mixing records from different cities or seasons. It also made it possible to connect each meteorological output directly to the corresponding Landsat scene used in the spatial analysis.

```
FILES = [
    {
        "city": "Wrocław",
        "date": "2024-07-26",
        "season": "Summer",
        "path": "Wroklaw_26july.nc",
    },
    ...
]
.
.
.
ds = xr.open_dataset(path)

print("\n----------------------------------------")
print(f"Processing: {item['city']} | {item['date']} |
{item['season']}")
print("Variables:", list(ds.data_vars))
print("Dimensions:", dict(ds.sizes))
```

Figure 3.15. Python file-organization structure used to link each ERA5-Land NetCDF file to its corresponding city, Landsat acquisition date, and season before variable extraction.

Extraction of meteorological variables.

For each NetCDF file, the required ERA5-Land variables were extracted in Python. Air temperature was taken from the 2 m temperature variable, while dewpoint temperature was taken from the 2 m dewpoint temperature variable. Wind speed was calculated from the 10 m u and v wind components. Incoming solar radiation was obtained from the surface solar radiation downwards variable. These extracted variables were used to calculate the acquisition-hour weather indicators reported later in Section 3.2.3.

```python
def get_variable(ds, name):
    """
    Safely extract a variable from an xarray dataset.
    """
    if name not in ds:
        raise KeyError(
            f"Variable '{name}' not found. Available variables:
{list(ds.data_vars)}"
        )
    return ds[name]

t2m = get_variable(ds, "t2m")

d2m = get_variable(ds, "d2m")

u10 = get_variable(ds, "u10")

v10 = get_variable(ds, "v10")

ssrd = get_variable(ds, "ssrd")
```

Figure 3.16. Python code used to verify and extract the ERA5-Land meteorological variables required for the acquisition-hour weather analysis.

Calculation of air temperature, wind speed, and relative humidity.

The extracted variables were converted into standard meteorological indicators. Air temperature was converted from Kelvin to degrees Celsius. Wind speed was calculated as the magnitude of the wind vector using the 10 m u and v components. Relative humidity was calculated from 2 m air temperature and 2 m dewpoint temperature using the Magnus approximation.

This produced air temperature, wind speed, and relative humidity values for each selected city-date combination.

```python
def relative_humidity_from_temperature_dewpoint(t2m_k, d2m_k):

    t_c = t2m_k - 273.15
    td_c = d2m_k - 273.15

    es = 6.112 * np.exp((17.67 * t_c) / (t_c + 243.5))
    e = 6.112 * np.exp((17.67 * td_c) / (td_c + 243.5))

    rh = 100.0 * (e / es)
    return np.clip(rh, 0, 100)
  .
  .
  .
t2m_c = t2m - 273.15
wind_speed = np.sqrt((u10 ** 2) + (v10 ** 2))
rh = relative_humidity_from_temperature_dewpoint(t2m, d2m)
```

Figure 3.17. Python calculations used to convert air temperature from Kelvin to Celsius, derive wind speed from u and v wind components, and calculate relative humidity from air temperature and dewpoint temperature.

Conversion of accumulated SSRD to mean hourly solar irradiance.

ERA5-Land does not give surface solar radiation downwards (SSRD) as instantaneous irradiance. It is stored as accumulated radiative energy, in J/m2. The Python script thus converted SSRD to mean hourly solar irradiance .

The conversion was performed by subtracting the hourly accumulated values for consecutive hours and dividing the difference by 3600 seconds.The final values were thus expressed in W/m 2 and are the mean solar irradiance over the chosen hourly interval.

```
def convert_ssrd_to_hourly_irradiance(ssrd):
    """
    Convert ERA5 / ERA5-Land surface solar radiation downwards
(SSRD)
    from accumulated energy (J/m²) to mean hourly solar
irradiance (W/m²).
    """
    time_dim = get_time_dimension(ssrd)

    ssrd = ssrd.sortby(time_dim)

    if ssrd.sizes[time_dim] > 1:
        ssrd_hourly_jm2 = ssrd.diff(dim=time_dim)

        ssrd_hourly_jm2 = ssrd_hourly_jm2.assign_coords(
            {time_dim: ssrd[time_dim].values[1:]}
        )

        solar_wm2 = ssrd_hourly_jm2 / 3600.0

    else:
        solar_wm2 = ssrd / 3600.0

    solar_wm2 = solar_wm2.where(solar_wm2 >= 0)

    return solar_wm2
```

Figure 3.18. Python function used to convert ERA5-Land accumulated surface solar radiation downwards from J/m² into mean hourly solar irradiance in W/m² by differencing consecutive hourly accumulation steps.

Temporal matching with the Landsat acquisition period.

After the SSRD conversion, the script selected the meteorological values corresponding to the Landsat acquisition period. In the processed ERA5-Land files, the timestamp after differencing represents the end of the hourly accumulation interval. For example, the 11:00 UTC value corresponds to the 10:00–11:00 UTC interval.

For this thesis, the 10:00–11:00 UTC interval was used as the closest hourly atmospheric reference to the Landsat acquisition period. The resulting values were used to describe the weather context of each selected scene. They were not used to calculate PET, UTCI, PMV, or any other outdoor thermal comfort index.

3.2 Candidate variables

The variables used in this thesis were selected to describe the main surface, morphology, and weather conditions related to SUHI analysis. Their selection follows the literature reviewed in Chapter 1 and the processing workflow described in Section 3.1 [67]. The same variable logic was applied to Wrocław and Turin so that the two case studies could be compared with a consistent structure.

The variables were not used in the same way throughout the thesis. Some variables were used as active predictors in the warm-season Random Forest models. Other variables were used only to describe the meteorological context of each Landsat acquisition or to support the seasonal spatial assessment of thermal anomalies. This distinction is important because the thesis combines spatial assessment and predictive modeling, but these two parts do not use all variables in the same role.

The variables were grouped into four categories, namely land-cover variables, urban geo-morphology variables, climate and weather-related variables, and SUHI evaluation variables. In the Random Forest phase, the target variable was SUHII. The land-cover and morphology variables selected were the active predictors, i.e. Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR and S/V Ratio. SUHII was derived from LST and hence LST was not used as a predictor as it would have introduced target leakage.

Climate and weather related variables were retained as contextual information. They were used to describe the atmospheric conditions for the selected satellite acquisition dates but not for the calculation of outdoor thermal comfort indices, such as PET, UTCI, or PMV. SUHI evaluation variables such as LST, SUHI, SUHII, UTFVI, and thermal hotspots were used mainly for the multi-seasonal spatial assessment and interpretation.

The final comparative modeling framework did not include human activity as a direct predictor. Population data were available only for one of the cities, Turin, but not for Wrocław, at a similar spatial scale. Providing population density for just one city would have reduced comparability between the two models. Therefore, possible effects of human activity were only indirectly considered, by means of built-environment proxies, such as NDBI, BCR, and S/V Ratio. Future studies could use population density, traffic intensity, energy-use data or nighttime lights if comparable datasets are available for both cities.

Table 3.3. Categorization and role of variables in the thesis workflow.

Category	Variables	Role in modeling
Land cover variables	Land use/land cover, NDVI, PVI, NDWI, NDMI, Albedo, Emissivity, NDBI	Describe vegetation, moisture, built-up intensity, and radiative surface properties. The selected index-based variables were used as active Random Forest predictors.

[67] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

Urban geo-morphology variables	SVF, BCR, S/V Ratio	Represent sky openness, building coverage, and urban compactness. These variables were used as active Random Forest predictors.
Climate and weather-related variables	Relative humidity, wind speed, mean hourly solar irradiance, air temperature	Provide atmospheric context for the selected Landsat acquisition dates. These variables were not used as thermal comfort indices.
SUHI evaluation variables	LST, SUHII, SUHI, UTFVI, thermal hotspots	Support the multi-seasonal spatial assessment and interpretation of thermal anomalies. SUHII was also used as the Random Forest target variable.

3.2.1 Land cover variables

Land cover variables were included because they describe the surface conditions that are most directly related to LST and SUHII. In urban areas, the replacement of natural and permeable surfaces with artificial materials changes the surface energy balance by reducing evapotranspiration, increasing heat storage, and modifying solar reflectance [68]. For this reason, vegetation, moisture, built-up intensity, and radiative surface properties were considered in the variable selection [69].

The thesis approached land cover in two ways. First, land use/land cover information offered a broad description of the physical surface context allowing for the differentiation of built-up areas, vegetation, open space and other surface types [70]. Secondly, the main land-cover variables in the modelling dataset were derived from Landsat imagery in the form of spectral and radiative indices. These index-based variables were more suitable for the Random Forest analysis, as they provided continuous numerical values at the scale of the census section after zonal aggregation.

Vegetation conditions were represented by the Normalized Difference Vegetation Index (NDVI) and the Proportion of Vegetation Index (PVI). NDVI was used to describe vegetation presence and greenness, while PVI was used to estimate the proportional vegetation component derived from NDVI [71]. These variables are relevant because vegetated surfaces can reduce surface heating through shading and evapotranspiration.

---

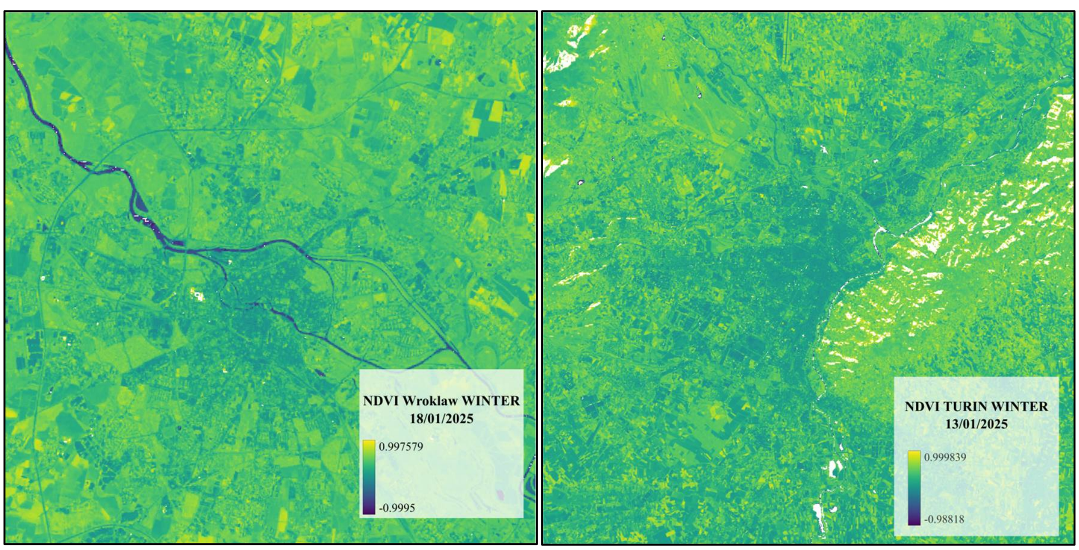
[68] Zhou et al., "Satellite Remote Sensing of Surface Urban Heat Islands."
[69] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[70] Sudarshana Sinha and Ankhi Banerjee, "A Systematic Review of Land Use and Land Cover Changes and Their Influence on Urban Heat Island Dynamics," *Discover Geoscience* 3, no. 1 (2025): 253, https://doi.org/10.1007/s44288-025-00359-4.
[71] Chanpichaigosol et al., "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

Moisture-related surface conditions were represented by the Normalized Difference Water Index (NDWI) and the Normalized Difference Moisture Index (NDMI). NDWI was used to identify water-related surface signals, while NDMI described surface and vegetation moisture conditions [72]. These variables were included because water and moisture can affect the surface energy balance and the cooling capacity of urban areas.

Built-up intensity was represented by the Normalized Difference Built-up Index (NDBI). This index was used to identify areas with a stronger presence of artificial and impervious surfaces, which are often associated with higher land surface temperatures and reduced evaporative cooling [73].

Radiative surface behavior was represented by Albedo and Emissivity. Albedo describes the fraction of incoming solar radiation reflected by the surface, while Emissivity describes the ability of the surface to emit thermal radiation [74]. These variables were included because different surface materials absorb, reflect, and emit heat in different ways.

Together, these variables describe the main surface properties used to explain SUHII in the two case studies: vegetation, moisture, built-up intensity, surface reflectivity, and thermal emissive behavior. Selected spatial examples for Wrocław and Turin are shown in Figures 3.19 to 3.21. The supplementary appendix focuses only on transitional-season SUHII and UTFVI maps, rather than repeating the full set of land-cover raster outputs.



[72] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania."
[73] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Xi et al., "Seasonal Surface Urban Heat Island Analysis Based on Local Climate Zones."
[74] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

Figure 3.19. NDVI distribution derived in Google Earth Engine for the Wrocław winter representative scene in 2025, showing the spatial variation of vegetation presence and surface greenness.

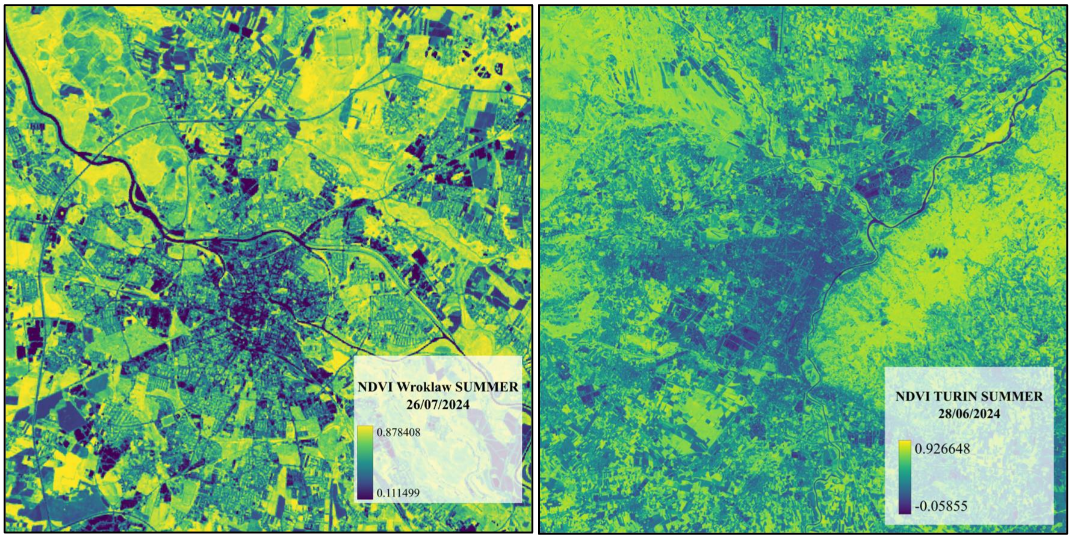
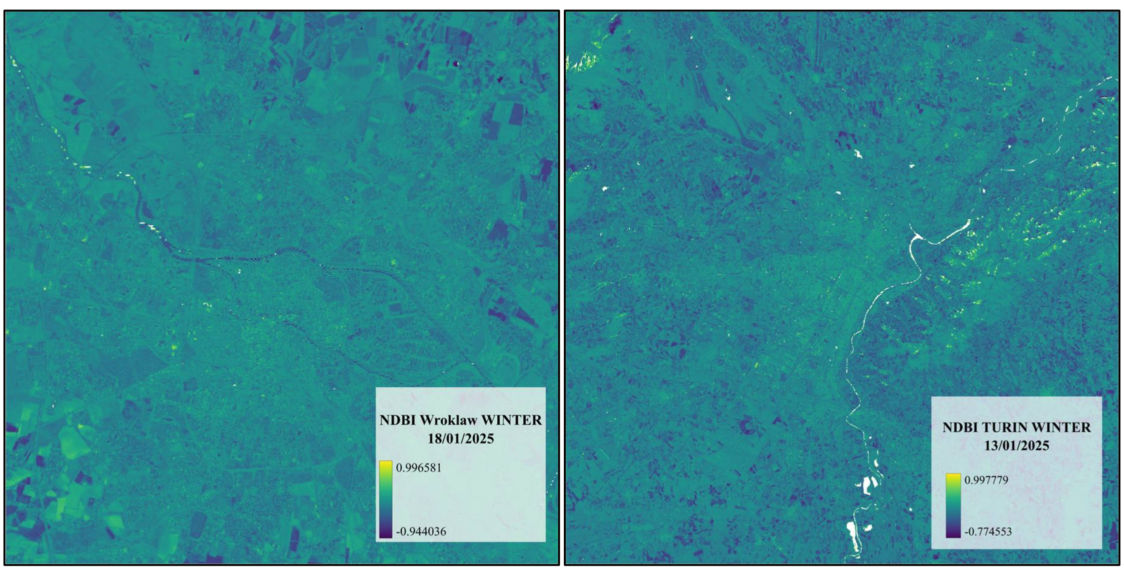


Figure 3.20. NDVI distribution derived in Google Earth Engine for the Wrocław summer representative scene in 2024, showing the stronger vegetation signal during the warm-season assessment.



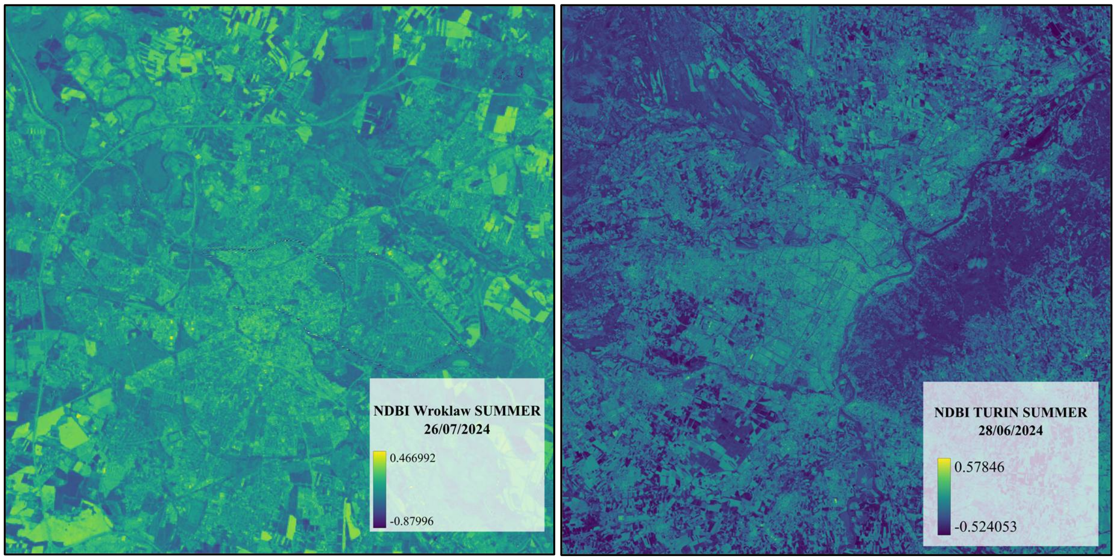


Figure 3.21. NDBI distribution derived in Google Earth Engine for the Wrocław winter and summer representative scenes, identifying the spatial concentration of built-up and impervious surfaces and its seasonal variation.

3.2.2 Urban geo-morphology variables

Urban geo-morphology variables were included to describe the built-form characteristics that cannot be captured by Landsat spectral indices alone. While land-cover indicators describe surface composition, morphology variables describe how the urban fabric is arranged in three dimensions. These characteristics can affect shading, sky openness, heat storage, and longwave radiation loss [75].

In this thesis, the morphology group included three variables: Sky View Factor (SVF), Building Coverage Ratio (BCR), and Surface-to-Volume Ratio (S/V Ratio). These variables were calculated or prepared at the census-section scale and were retained in the final Random Forest predictor set for both Wrocław and Turin.

SVF was used to represent sky openness and urban enclosure. It expresses the proportion of visible sky from a given location, with values ranging from 0 to 1 after normalization. In the original Turin dataset, SVF values were stored as percentages, so they were converted to dimensionless ratios before modeling. Lower SVF values indicate more enclosed urban spaces, where buildings or other elements obstruct the sky. This condition can reduce direct solar exposure in some daytime situations, but it can also limit longwave radiation loss and affect nighttime cooling [76].

[75] Oke, "The Energetic Basis of the Urban Heat Island"; Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity."
[76] Hidalgo García and Arco Díaz, "Modeling of the Urban Heat Island on Local Climatic Zones of a City Using Sentinel 3 Images"; Oke, "The Energetic Basis of the Urban Heat Island."

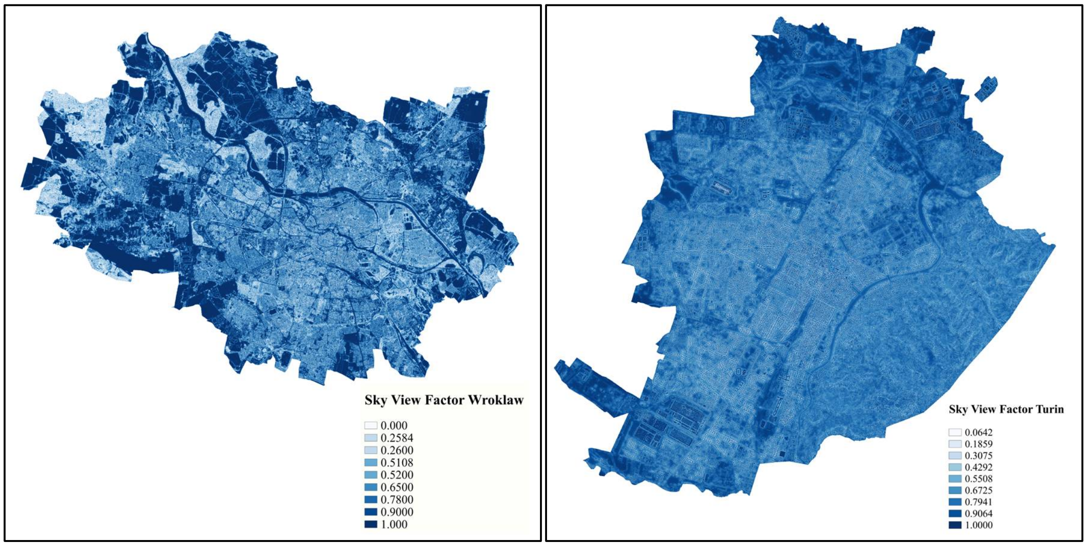


Figure 3.22. Sky View Factor (SVF) spatial distribution in Wrocław and Turin expressed as a dimensionless ratio from 0 to 1. Lower values indicate more enclosed urban canyons with less sky exposure, while higher values indicate more open areas with more sky visibility.

The horizontal compactness of the built environment was described by BCR. It was calculated as the ratio of the total building footprint area within a census section to the total area of that census section. This is a ratio and has a theoretical value between 0 and 1 .[77]. The Turin BCR values were re-computed and validated (to represent a valid building-footprint-to-census-area ratio) prior to the final modeling stage. After this correction, BCR was added as a predictor for both Turin and Wroclaw.

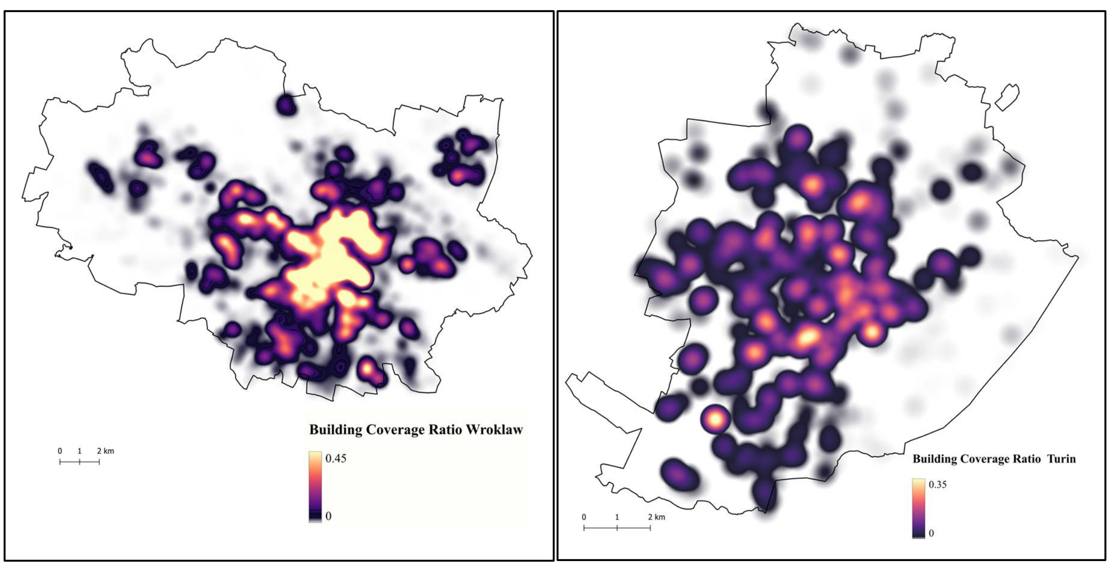


Figure 3.23. Spatial distribution of the Building Coverage Ratio (BCR) in Wroclaw and Turin, calculated as the ratio of total building footprint area to census-unit area Higher scores mean a larger share of built-up coverage and a more compact horizontal urban fabric.

The S/V Ratio was used to express the relationship between exposed building surface and building volume. This variable provides information on the compactness of the geometry of the building and the

[77] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

amount of surface possibly involved in heat exchange [78]. In general, higher S/V values indicate more exposed or less compact built forms, and lower values indicate more compact forms. Building Volume was used during the calculation of S/V Ratio, but it was not included as a separate final predictor in the Random Forest models.

Together, SVF, BCR, and S/V Ratio provide the morphology component of the modeling dataset. They were used alongside the Landsat-derived predictors so that the Random Forest models could account for both surface properties and selected built-form conditions.

3.2.3 Climate and weather-related variables

Climate and weather-related variables were included to describe the atmospheric conditions during the selected Landsat acquisition dates. Surface temperature patterns are mainly analyzed in this thesis through land cover, radiative properties, and urban morphology, but the weather conditions at the time of image acquisition also affect how strongly surfaces heat up. For this reason, the meteorological variables were used as contextual information for interpreting the seasonal thermal maps and the differences between Wrocław and Turin [79].

The variables selected for this part were air temperature, wind speed, mean hourly solar irradiance, and relative humidity. Air temperature describes the near-surface atmospheric background during the satellite acquisition. Wind speed gives a simple indication of potential ventilation and heat dispersion. Mean hourly solar irradiance represents the incoming solar energy available for surface heating during the acquisition hour. Relative humidity helps describe the moisture condition of the atmosphere and supports the interpretation of possible latent heat exchange [80].

These variables were not used to calculate outdoor thermal comfort indices such as PET, UTCI, or PMV. They were also not included in the final Random Forest predictor set. Their role was to provide a comparable meteorological context for the selected scenes and to help explain why thermal conditions may differ between seasons or between the two case-study cities.

---

[78] Guglielmina Mutani and Valeria Todeschi, “Roof-Integrated Green Technologies, Energy Saving and Outdoor Thermal Comfort: Insights from a Case Study in Urban Environment,” *International Journal of Sustainable Development and Planning* 16, no. 1 (2021): 13–23, https://doi.org/10.18280/ijsdp.160102.

[79] Sirous Haashemi et al., “Seasonal Variations of the Surface Urban Heat Island in a Semi-Arid City,” *Remote Sensing* 8, no. 4 (2016): 352, https://doi.org/10.3390/rs8040352; Gabriel Yoshikazu Oukawa et al., “Fine-Scale Modeling of the Urban Heat Island: A Comparison of Multiple Linear Regression and Random Forest Approaches,” *Science of The Total Environment* 815 (April 2022): 152836, https://doi.org/10.1016/j.scitotenv.2021.152836.

[80] Guglielmina Mutani et al., “Improving Outdoor Thermal Comfort in Built Environment Assessing the Impact of Urban Form and Vegetation,” *International Journal of Heat and Technology* 40, no. 1 (2022): 23–31, https://doi.org/10.18280/ijht.400104.

The meteorological data were downloaded from the Copernicus Climate Data Store from ERA5-Land hourly reanalysis data as described in Section 3.1.3. ERA5-Land was selected because it provides gridded data with the same structure for Wrocław and Turin. There was a meteorological station near Wroclaw, but a single station would not have provided the same spatially consistent reference for the two cities. To ensure comparability of the meteorological source in the two case studies, reanalysis data were used.

For each selected Landsat date, an ERA5-Land NetCDF file was matched and processed using Python. The air temperature was derived from the 2m temperature variable and converted from Kelvin to degrees Celsius. Wind speed was calculated from the 10 m zonal and meridional wind components as the magnitude of the wind vector. Relative humidity was computed from the 2 m air temperature and the 2 m dewpoint temperature using the Magnus approximation. Incoming solar radiation was taken from the surface solar radiation downwards (SSRD) variable.

Because SSRD is stored in ERA5-Land as accumulated radiative energy in J/m², it was converted to mean hourly solar irradiance. This was done by subtracting consecutive hourly accumulation values and dividing the difference by 3600 seconds. The resulting values are expressed in W/m² and represent the mean solar irradiance during the Landsat acquisition-hour interval.

The 10:00–11:00 UTC interval was used as the closest hourly atmospheric reference to the Landsat acquisition period. The resulting values are summarized in Table 3.4.

Table 3.4. Meteorological conditions during the Landsat acquisition hour for the selected case-study dates.

City	Date	Season	Air temperature at acquisition (°C)	Wind speed (m/s)	Mean solar irradiance during acquisition hour (W/m²)	Relative humidity (%)
Wrocław	2024-07-26	Summer	32.64	4.07	811.51	31.20
	2024-09-20	Autumn	18.65	3.64	613.42	55.78
	2025-01-18	Winter	3.48	0.89	252.13	79.26
	2025-05-05	Spring	11.02	4.34	614.18	47.97
Turin	2024-06-28	Summer	27.16	1.26	873.87	66.96
	2024-11-10	Autumn	23.91	3.89	701.33	34.41
	2025-01-13	Winter	3.77	1.09	365.67	46.27
	2025-05-30	Spring	23.30	1.65	883.37	57.58

Before moving to the SUHI evaluation indicators, the main input datasets used in the thesis are summarized in Table 3.5. The table reports the type of dataset, the variables derived or extracted from it, its spatial or temporal resolution, the data provider, and its role in the analysis.

This summary is added to clarify the structure of data in the thesis. The workflow integrates several types of input data: Environmental and thermal variables derived from Landsat, ESA WorldCover land-cover data for the rural reference mask, GIS-based urban morphology variables, ERA5-Land meteorological data, and census-section polygons used for zonal aggregation and modeling.

The Landsat-derived variables were originally generated from imagery with a native spatial resolution of 30 m. During GIS preparation, the raster layers were harmonized to a common 10 m working grid, so that the spatial structure used for the overlay with ESA WorldCover and morphology-related GIS layers was consistent. This resampling step was applied only for the spatial alignment and the zonal statistics. It did not add any new spatial information nor did it increase the original detail of the Landsat imagery.

Table 3.5. Main input datasets used for the Wrocław and Turin SUHI analysis, including native resolution and working resolution after spatial harmonization.

Dataset	**Variables/outputs used in the thesis**	**Native resolution / working scale**	**Source**	**Link/access**	**Reference/role in the thesis**
Landsat 8/9 Collection 2 Level 2	NDVI, NDBI, NDWI, NDMI, Albedo, PVI, Emissivity, LST, SUHII, SUHI-related raster layers	Native spatial resolution: 30 m; harmonized/reprojected to a 10 m working grid for GIS integration and zonal statistics	USGS / NASA, accessed through Google Earth Engine	USGS Landsat Collection 2 Level-2 Science Products; Google Earth Engine Landsat Data Catalog	Main remote-sensing source for surface reflectance, thermal information, spectral indices, LST, and SUHII calculation
Google Earth Engine platform	Image filtering, cloud-cover filtering, scale-factor application, raster generation, export to GeoTIFF	Cloud-based processing environment	Google Earth Engine	Google Earth Engine Data Catalog (Google for Developers)	Operational environment used for preprocessing, index calculation, LST/SUHII derivation, and raster export

ESA WorldCover 2021 v200	Rural reference mask for SUHI calculation; retained classes: tree cover, shrubland, grassland, cropland, bare/sparse vegetation, herbaceous wetland	10 m	European Space Agency	ESA WorldCover 2021 v200 / GEE asset ESA/WorldCover/v200 (esa-worldcover.org)	Used to separate rural reference areas from built-up areas for SUHI calculation
ERA5-Land hourly reanalysis	2 m air temperature, 2 m dewpoint temperature, 10 m wind components, wind speed, relative humidity, surface solar radiation downwards, mean hourly solar irradiance	Hourly; approximately 9 km grid	Copernicus Climate Data Store / ECMWF	ERA5-Land hourly data from 1950 to present (Climate Data Store)	Meteorological context for each Landsat acquisition date; processed in Python
Census-section / urban statistical units	Spatial aggregation units for zonal statistics and Random Forest modeling	Polygon units; city-specific census/statistical scale	Local GIS working datasets/thesis GeoPackages	Final Wrocław and Turin GeoPackages to be uploaded to the thesis folder	Used to aggregate raster variables from pixel scale to polygon-based modeling units
Building footprint and urban morphology layers	Building footprint area, built-up coverage, BCR, building geometry information for S/V Ratio	Vector layers; aggregated to census-section scale	Local GIS working datasets/thesis GeoPackages	Final Wrocław and Turin GeoPackages to be uploaded to the thesis folder	Used to calculate BCR and S/V Ratio, and to describe urban compactness

Sky View Factor layer	SVF	Raster or polygon-derived values; converted to a dimensionless ratio from 0 to 1	GIS-derived morphology layer	Final Wrocław and Turin GeoPackages to be uploaded to the thesis folder	Used to represent sky openness, canyon enclosure, and potential longwave-radiation trapping
QGIS zonal statistics outputs	Mean values of Landsat-derived variables per census section	Census-section scale	Generated in QGIS	Final modeling CSV / GeoPackage outputs	Converted raster layers into structured polygon-based datasets for Python and Random Forest modeling
Final modeling dataset	Predictor variables and SUHII target for Wrocław and Turin	Census-section scale	Generated from GEE, QGIS, and Python workflow	Final CSV / GeoPackage outputs	Input table for Random Forest regression and feature-importance analysis

3.2.4 SUHI evaluation

This section explains the thermal indicators used to evaluate surface heat patterns in Wrocław and Turin. The analysis included Land Surface Temperature (LST), Surface Urban Heat Island Intensity (SUHII), Surface Urban Heat Island magnitude (SUHI), the Urban Thermal Field Variance Index (UTFVI), and thermal hotspot mapping. These indicators were used mainly for the multi-seasonal spatial assessment. SUHII was also used as the target variable in the warm-season Random Forest models.

LST was derived from Landsat thermal data in Google Earth Engine after emissivity correction. It is the surface temperature at the moment of the satellite acquisition and was the initial thermal layer for all the indicators presented in this section. However, LST was not used as a predictor in the Random Forest models since SUHII is derived from LST . So, including LST in the predictor set would be target leakage.

SUHI (I was the standardized measure of the relative surface thermal anomaly for each Landsat scene. This enabled the comparison of warmer and cooler areas within the same study area without relying solely on absolute LST values that are strongly affected by season and acquisition conditions. The formula used was:

$$SUHII\_{i} = (LST\_{i} - LST_{mean}) / LST_{std}$$

where $SUHII\_{i}$ is the standardized Surface Urban Heat Island Intensity for pixel or spatial unit i; LST_{i} is the land surface temperature of pixel or spatial unit i; LST_{mean} is the mean land surface temperature of the study area for the same Landsat scene; and LST_{std} is the standard deviation of land surface temperature within the same study area and acquisition date

Positive SUHII values indicate surfaces warmer than the scene-specific mean LST, while negative values indicate relatively cooler surfaces. In this thesis, SUHII was used in two ways. First, it supported the multi-seasonal spatial assessment of thermal anomalies. Second, the warm-season SUHII layer was used as the target variable for the Random Forest regression in Chapter 4.

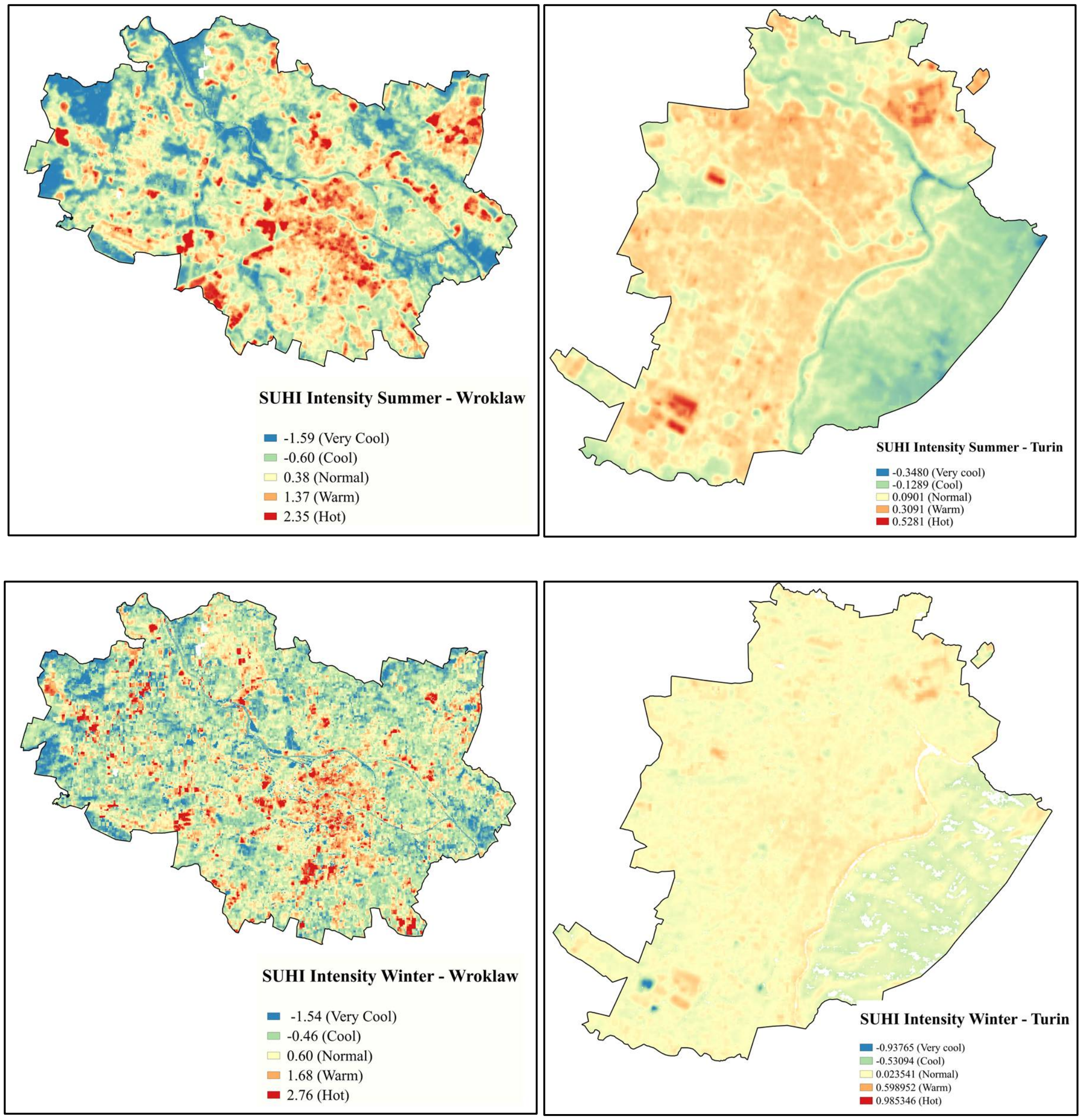


Figure 3.24. Standardized Surface Urban Heat Island Intensity (SUHII) maps for representative Landsat scenes in Wrocław and Turin. Positive values indicate surfaces warmer than the scene-specific mean LST, while negative values indicate relatively cooler surfaces.

In addition to SUHII, a SUHI raster was calculated to express the surface temperature difference from a rural reference condition. This indicator was used to map the urban–rural surface thermal contrast more directly. SUHI was calculated as:

$$SUHI\_i = LST\_i - mean(LST\_rural)$$

where *SUHI_i* is the Surface Urban Heat Island value for pixel *i*; *LST_i* is the land surface temperature of pixel *i* within the study area; and *mean(LST_rural)* is the mean land surface temperature of the rural reference pixels for the same Landsat acquisition date

The mask for the rural reference was created from the ESA WorldCover 2021 v200 land cover product in Google Earth Engine. The WorldCover "Map" band was clipped to the AOI of each case-study city, and used to identify rural reference pixels. The classes retained were Tree cover (10), Shrubland (20), Grassland (30), Cropland (40), Bare/sparse vegetation (60), and Herbaceous wetland (90). (50) Urbanized surfaces were excluded because they are built-up areas. Permanent water bodies (80) were also excluded since their thermal behavior is different from vegetated or agricultural rural surfaces and might bias the rural LST reference.

The LST raster was masked for each Landsat acquisition date using this rural reference layer. Then, the mean LST of the retained rural pixels was calculated at 30 m scale to be consistent with the Landsat thermal output . The SUHI raster was derived by subtracting the rural mean value from each LST pixel. The same ESA WorldCover based procedure was applied to Wrocław and Turin to ensure comparability of the SUHI calculation between the two case studies.

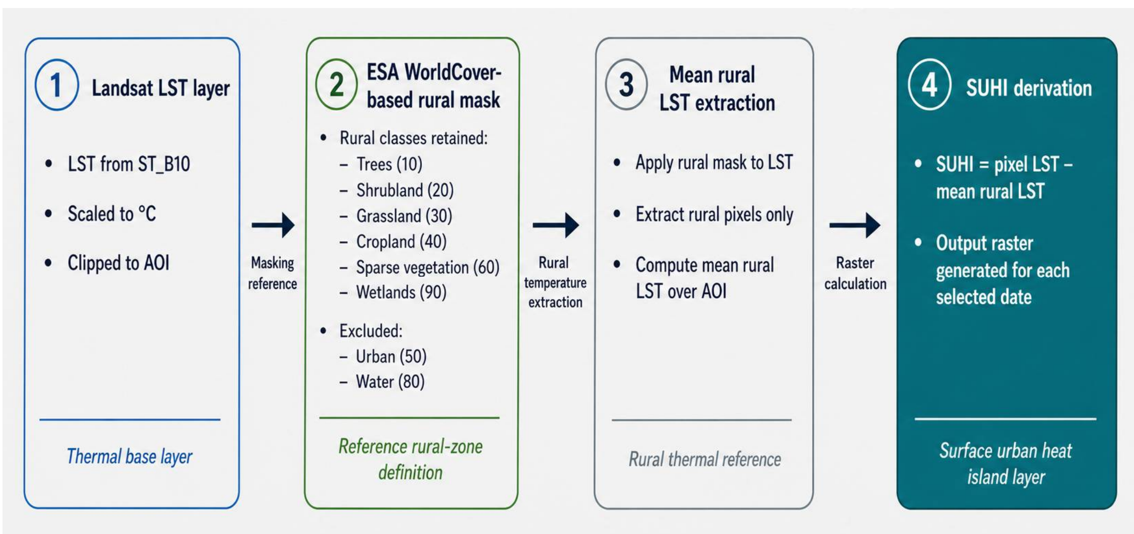


Figure 3.25. Methodological workflow used in Google Earth Engine to define the rural reference mask from ESA WorldCover 2021 v200 classes and derive the pixel-based SUHI raster from Landsat LST.

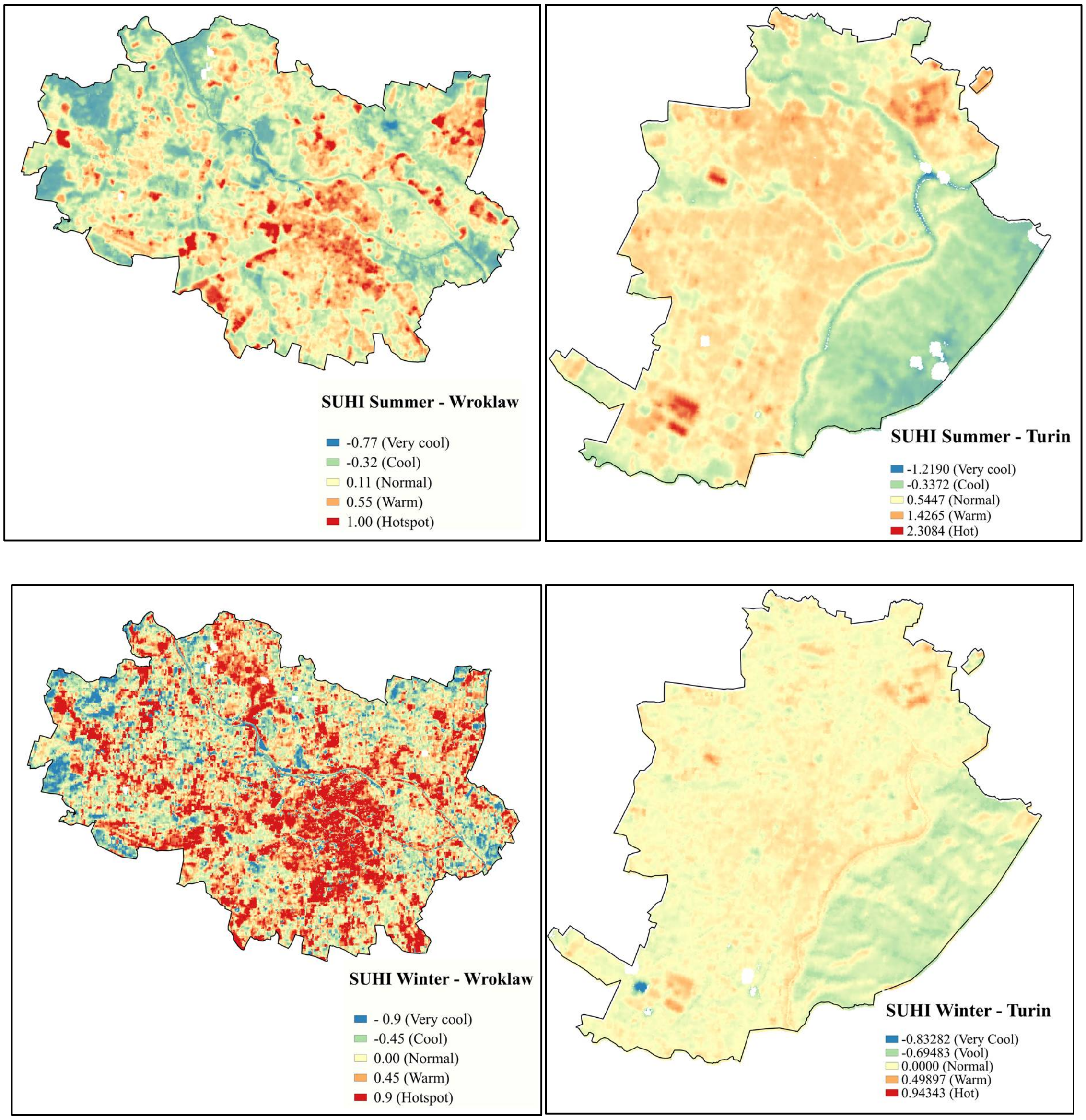


Figure 3.26. Representative Surface Urban Heat Island (SUHI) maps for Wrocław and Turin, calculated as the pixel-based LST difference from the mean rural reference temperature for each Landsat scene.

The Urban Thermal Field Variance Index (UTFVI) was calculated in QGIS from the LST layer. UTFVI was used as an additional indicator to classify relative ecological thermal stress based on surface temperature variation [81]. The formula used was:

$$UTFVI\_i = (LST\_i - mean\ (LST))\ /\ mean\ (LST)$$

[81] Kara et al., "Multi-Index Assessment of Surface Urban Heat Island (SUHI) Dynamics in Samsun Using Google Earth Engine."

where *UTFVI_i* is the Urban Thermal Field Variance Index for pixel *i*; *LST_i* is the land surface temperature of pixel *i*; and *mean (LST)* is the mean land surface temperature of the study area for the same Landsat scene.

Higher UTFVI values indicate stronger relative thermal stress compared with the mean condition of the same scene. The resulting UTFVI layers were classified into ecological thermal stress categories ranging from "None" to "Strongest" according to the adopted threshold structure [82]. These maps were used to support the interpretation of thermally stressed areas. They were not used to calculate PET, UTCI, PMV, or any other outdoor thermal comfort index.

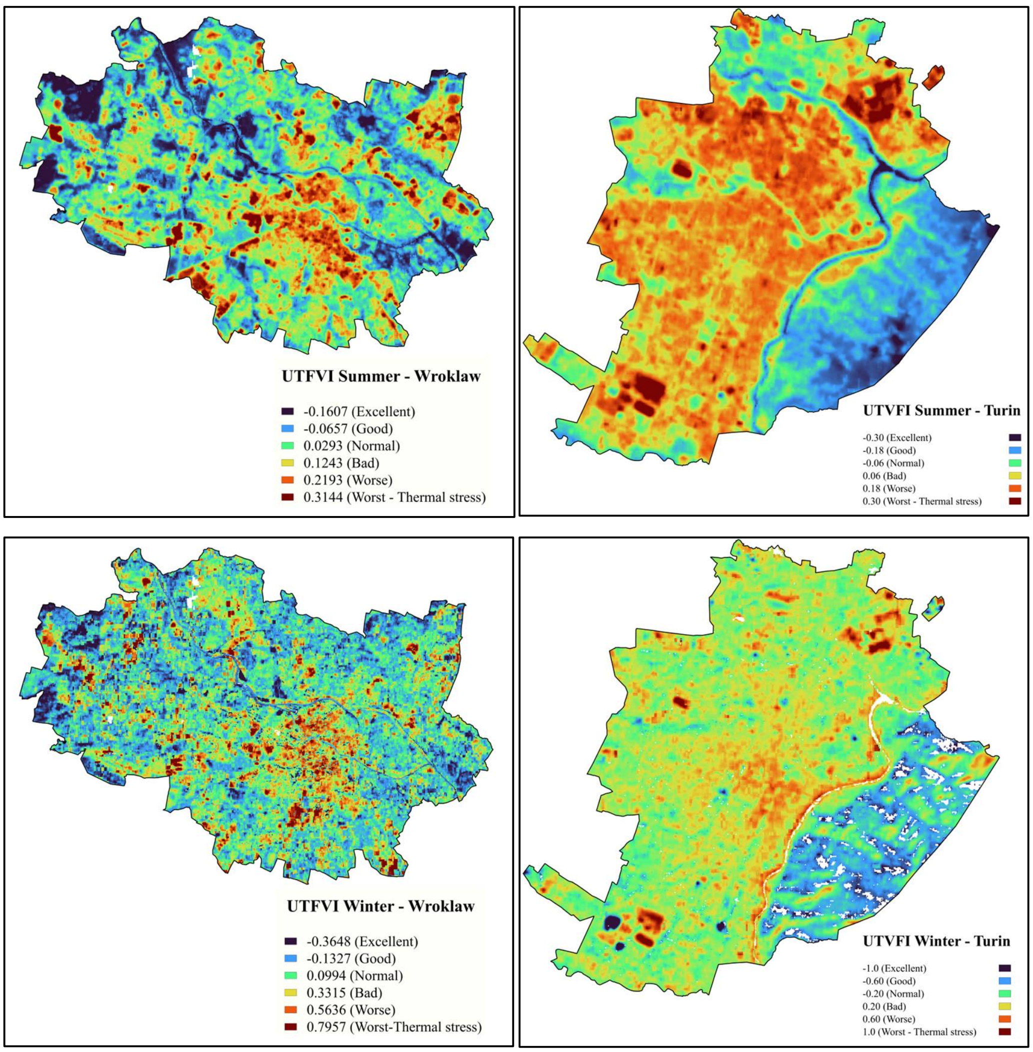


Figure 3.27. Urban Thermal Field Variance Index (UTFVI) distribution in Wrocław and Turin, classified into ecological thermal stress levels from none to strongest according to the relative LST anomaly of each scene.

[82] Kara et al., "Multi-Index Assessment of Surface Urban Heat Island (SUHI) Dynamics in Samsun Using Google Earth Engine."

Thermal hotspot analysis was then performed in QGIS by combining SUHI and UTFVI. A pixel or spatial unit was classified as a hotspot when it was warmer than the rural reference condition and also belonged to the high UTFVI stress classes. The logical rule was:

$$Hotspot\_i = 1 \text{ if } (SUHI\_i > 0) \text{ AND } (UTFVI\_i \geq UTFVI\_high), \text{ otherwise } Hotspot\_i = 0$$

where *Hotspot_i* identifies whether pixel or spatial unit *i* is classified as a thermal hotspot; *SUHI_i* > 0 indicates that the location is warmer than the rural-reference condition; and *UTFVI_i* ≥ *UTFVI_*high indicates that the same location belongs to the high ecological thermal stress classes. In this thesis, *UTFVI_high* refers to the "Strong", "Stronger", and "Strongest" UTFVI classes

This procedure was used to identify areas where positive urban–rural thermal contrast and high relative thermal stress occurred at the same time. The hotspot maps were prepared for the selected seasonal scenes and then used to support the planning interpretation in Chapter 5. In the final step, the hotspot layers from the four selected dates were combined to produce a composite annual hotspot map showing the areas where thermal stress appeared more recurrently across the seasonal assessment.

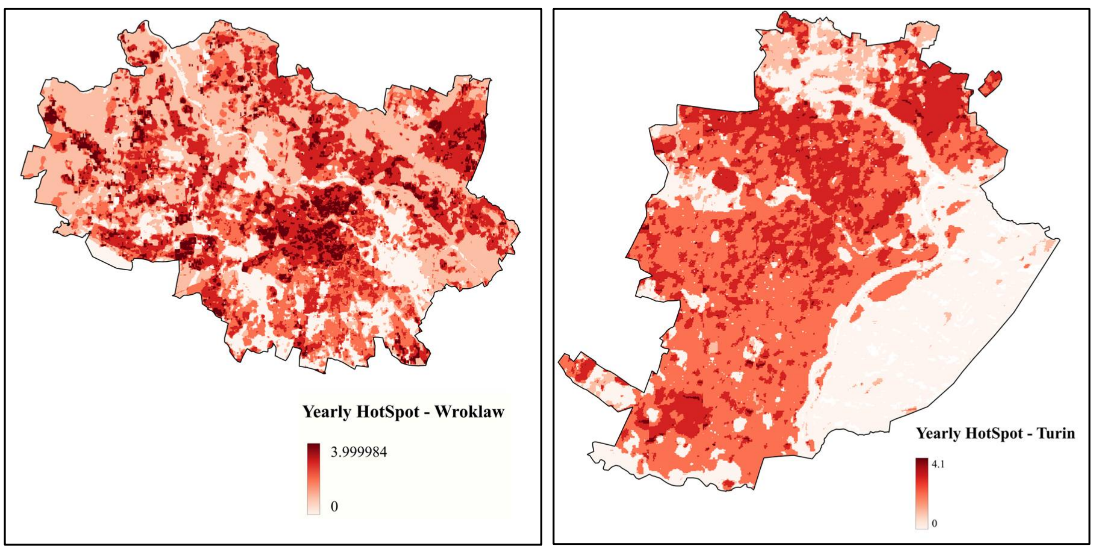


Figure 3.28. Spatial distribution of thermal hotspots in Wrocław and Turin, identified through the logical intersection of positive SUHI anomalies and high UTFVI ecological stress classes.

Together, these outputs provided the thermal basis for the later spatial interpretation. SUHII was used as the standardized anomaly layer and became the target variable for the Random Forest models in Chapter 4. LST, SUHI, UTFVI, and hotspot recurrence were used to describe the seasonal distribution of surface heat, rural-reference thermal contrast, relative thermal stress, and recurrent hotspot areas.

These layers were then used in Chapter 5 to support the planning discussion of possible mitigation measures, without treating them as simulated before-and-after intervention results [83].

[83] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

4. Modelling: Variables Influencing Surface Urban Heat Island Effect

Chapter 3 described how the remote-sensing, morphology, meteorological, and thermal assessment layers were prepared. The present chapter focuses on the predictive modeling of Surface Urban Heat Island Intensity (SUHII). The modeling dataset was prepared from the census-section attribute table produced in QGIS through zonal statistics. Each spatial unit contained the mean values of the selected raster-derived indicators and morphology variables, and the table was then exported as a CSV file for Python-based machine-learning analysis.

The purpose of this chapter is to test how well the selected predictors explain warm-season SUHII in Wrocław and Turin, and to identify which variables have the strongest role in the models. The same modeling logic was first applied to Wrocław and then repeated for Turin. Random Forest regression was selected because it can model non-linear relationships between SUHII and variables related to vegetation, moisture, built-up intensity, radiative surface behavior, and urban morphology [84].

Although Chapter 3 included four representative seasonal scenes for each city, the final Random Forest models were developed only for the warm-season representative scenes. This choice was made because the warm season provided the clearest surface thermal contrast and was therefore more suitable for identifying the spatial drivers of SUHII [85]. The other seasonal layers were retained for the multi-seasonal spatial assessment and planning interpretation, not as separate predictive models.

SUHII was used as the target variable. LST was not included in the predictor set because SUHII was calculated from LST; using LST as a predictor would therefore introduce target leakage. The final predictor set used consistently for both cities included Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio.

The modeling workflow is summarized in Figure 4.1. It includes data preparation, hyperparameter tuning, model training and validation, feature-importance analysis, and spatial prediction in ArcGIS Pro.

[84] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Oukawa et al., "Fine-Scale Modeling of the Urban Heat Island."

[85] Esposito et al., "A Multi-City Statistical Modelling of Surface Urban Heat Island."

Figure 4.1. Workflow for warm-season SUHII modeling using Random Forest regression, from data preparation and hyperparameter tuning to validation, feature-importance analysis, and spatial prediction.

4.1 Hyperparameter Tuning

Before training the final Random Forest models, the census-based CSV table exported from QGIS was prepared in Python. This table contained one record for each census-section unit and the mean values of the selected raster-derived and morphology variables. Only the warm-season variables were used for the predictive modeling stage, while the other seasonal layers were retained for spatial assessment and interpretation.

The Python workflow included data cleaning, variable selection, train-test splitting, hyperparameter tuning, model evaluation, feature-importance extraction, and preparation of the outputs used for spatial prediction. The general structure of this workflow is shown in Figure 4.2.

```python
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import
RandomizedSearchCV
rf = RandomForestRegressor(random_state=42)

param_dist = {
    "n_estimators": [200, 500, 1000],
    "max_depth": [10, 20, 30, 40, None],
    "max_features": ["sqrt", "log2"],
    "min_samples_split": [2, 5, 10],
    "min_samples_leaf": [1, 2, 4]
}

random_search = RandomizedSearchCV(
    estimator=rf,
    param_distributions=param_dist,
    n_iter=50,
    cv=5,
    scoring="r2",
    random_state=42,
    n_jobs=-1,
    verbose=1
)

random_search.fit(X_train, y_train)

print(random_search.best_params_)
print(random_search.best_score_)
```

Figure 4.2. Python-based workflow used for Random Forest data preparation, hyperparameter tuning, model evaluation, feature-importance extraction, and spatial prediction support.

SUHII was selected as the target variable. The active predictors were Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio. LST was excluded from the predictor set because SUHII was derived from LST, and including it would have introduced target leakage.

BCR was recalculated for the Turin dataset before the final modelling step to consider the ratio between building footprint area and census section area. Following this correction, BCR values were compared with the theoretical range of 0 to 1. BCR and S/V Ratio were then retained in the Turin model, obtaining the same final set of ten predictors as for Wrocław.

Hyper parameter tuning was done using five fold cross validation. During the tuning process, some Random Forest parameters were explored such as the number of trees, the maximum depth of the trees, the number of variables to be considered at each split, the minimum number of samples required to split a node and the minimum number of samples required at a node leaf. The execution of this step in Python is shown in Figure 4.3.

```
Fitting 5 folds for each of 50 candidates, totalling 250 fits
Best hyperparameters for Wroklaw:

{'n_estimators': 200, 'min_samples_split': 2,
'min_samples_leaf': 2, 'max_features': None, 'max_depth': 40}

Best cross-validation R2 score:
0.7441
```

```
Fitting 5 folds for each of 50 candidates, totalling 250 fits
Best hyperparameters for Turin:
{'n_estimators': 1000, 'min_samples_split': 2,
'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': 10}

Best cross-validation R2 score:
0.7737
```

Figure 4.3. Python execution of Random Forest hyperparameter tuning and five-fold cross-validation for the SUHII models.

The Wrocław and Turin models were tuned separately because the two datasets had different numbers of spatial units and different predictor distributions. Using independent tuning avoided forcing the same parameter configuration onto both case studies. The selected hyperparameter settings were then used for final model training and test-set evaluation.

The optimized parameters and cross-validation results are reported in Table 4.1 for Wrocław and Table 4.2 for Turin.

Table 4.1. Optimized Random Forest hyperparameter settings for the Wrocław SUHII model.

Hyperparameter	Selected value	Modelling role
n_estimators	200	Number of decision trees in the forest
max_depth	40	Maximum depth allowed for each decision tree
max_features	None	All predictor variables are considered for the best split
min_samples_split	2	Min. samples required to split an internal node
min_samples_leaf	2	Min. samples required at a terminal leaf node
Best CV R^2	0.7441	Best cross-validation performance obtained during tuning

Table 4.2. Optimized Random Forest hyperparameter settings for the Turin SUHII model.

Hyperparameter	Selected value	Modelling role
n_estimators	1000	Number of decision trees in the forest
max_depth	30	Maximum depth allowed for each decision tree
max_features	log2	Number of predictor variables considered at each split
min_samples_split	15	Min. samples required to split an internal node
min_samples_leaf	1	Min. samples required at a terminal leaf node
Best CV R^2	0.7737	Best cross-validation performance obtained during tuning

4.2 Random Forest Regression Algorithm

After hyperparameter tuning, the final Random Forest regression models were built separately for Wrocław and Turin using the selected city-specific parameter settings. Random Forest was used because SUHII is influenced by several surface and morphology variables at the same time, and these relationships are not expected to be purely linear. This makes the method suitable for testing how vegetation, moisture, built-up intensity, radiative surface properties, and selected urban morphology indicators relate to SUHII [86].

Random Forest is an ensemble learning method based on multiple decision trees. Each tree is trained on a bootstrap sample of the dataset, and a subset of predictors is considered when the tree is split. For regression tasks, the final prediction is obtained by averaging the predictions from all trees. In this thesis, the model therefore produced a continuous predicted SUHII value for each census-section unit.

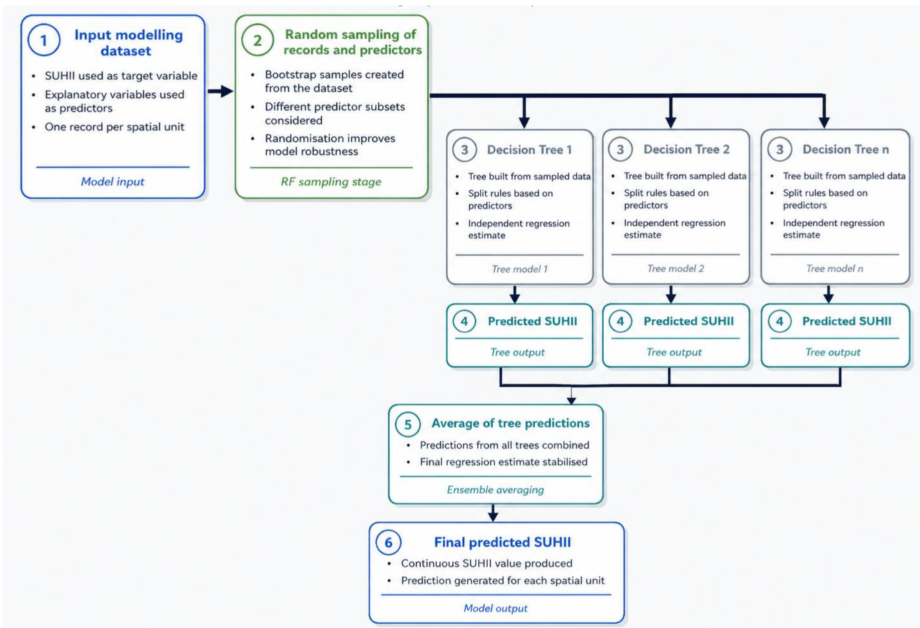


Figure 4.4. Simplified decision-tree logic within the Random Forest regression structure used for SUHII prediction.

Random Forest was applied as a regression model, not as a classification model. This distinction is important because SUHII is a continuous variable, not a categorical class. The model performance was therefore evaluated using regression metrics: the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) [87].

The final models used the same ten active predictors for both case studies: Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio. SUHII was the target variable. LST was excluded from the predictor set because SUHII was calculated from LST, and including LST would have introduced target leakage. The predictor and target structure is summarized in Table 4.3.

Table 4.3. Active predictors and target variable used in the Wrocław and Turin SUHII Random Forest models.

---

[86] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Oukawa et al., "Fine-Scale Modeling of the Urban Heat Island."

[87] Oukawa et al., "Fine-Scale Modeling of the Urban Heat Island."

Case Studies	Active predictors used in training	Target variable
Wrocław	Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, S/V Ratio	**SUHII**
Turin	Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, S/V Ratio	**SUHII**

The Random Forest models also produced feature-importance scores. These scores were used to identify which predictors contributed most to reducing model error [88]. In this thesis, feature importance was used as an explanatory support for interpreting SUHII drivers. However, these scores do not prove direct physical causality by themselves. For this reason, the feature-importance results were interpreted together with the spatial assessment described in Chapter 3 and the planning discussion developed in Chapter 5.

The modeling was carried out in Python for data cleaning, train-test splitting, hyperparameter tuning, model evaluation, residual diagnostics, and feature-importance extraction. The resulting predictions were then connected to the GIS workflow in ArcGIS Pro to produce the spatial SUHII prediction maps discussed later in this chapter.

4.3 Training, Validation, and Spatial Prediction of SUHII in the Case Studies

After the Random Forest parameters were selected, the final models were trained and evaluated separately for Wrocław and Turin. The purpose of this stage was to test how well the selected predictors could estimate warm-season Surface Urban Heat Island Intensity (SUHII) in each case study.

The modeling process was carried out in Python. For each city, the cleaned census-section dataset was split into training and testing subsets using an 80/20 ratio. The training subset was used to fit the Random Forest model, while the test subset was kept separate for the final evaluation. This separation was used to avoid reporting only an in-sample fit and to assess model performance on data not used during training [89].

As described in the previous sections, SUHII was used as the target variable. The active predictors were Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio. LST was not included in the predictor set because SUHII was derived from LST. The configuration of the training and testing datasets is summarized in Table 4.4.

Table 4.4. Configuration of training and testing datasets for the Wrocław and Turin SUHII models.

[88] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[89] Oukawa et al., "Fine-Scale Modeling of the Urban Heat Island."

Case Studies	Records	Active predictors	Variables retained only for reporting	Target variable	Training share	Testing share
Wrocław	645	10	None	SUHII	80%	20%
Turin	4688	10	None	SUHII	80%	20%

Model performance was evaluated on the independent test set. Three regression metrics were used: the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). R^2 describes the proportion of SUHII variance explained by the model, while RMSE and MAE describe the magnitude of prediction error [90]. The results are reported in Table 4.5.

Table 4.5. Performance metrics for the warm-season SUHII predictive models based on independent test sets.

Case study	Test R^2	Test RMSE	Test MAE	Best CV R^2
Wrocław	0.7576	0.3068	0.2400	0.7441
Turin	0.7841	0.3260	0.2485	0.7737

The Wrocław model reached a test R^2 of 0.7576, with an RMSE of 0.3068 and an MAE of 0.2400. The Turin model reached a test R^2 of 0.7841, with an RMSE of 0.3260 and an MAE of 0.2485. These values indicate that the selected predictors explained a large part of the warm-season SUHII variation in both case studies. The Turin model showed a slightly higher R^2, while the error values remained close to those obtained for Wrocław.

Diagnostic plots were prepared from the independent test-set predictions. These plots included observed versus predicted SUHII, residuals versus predicted SUHII, and the residual distribution. They were used to check the relation between observed and predicted values and to inspect whether the residuals showed visible bias or uneven error patterns.

[90] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania."

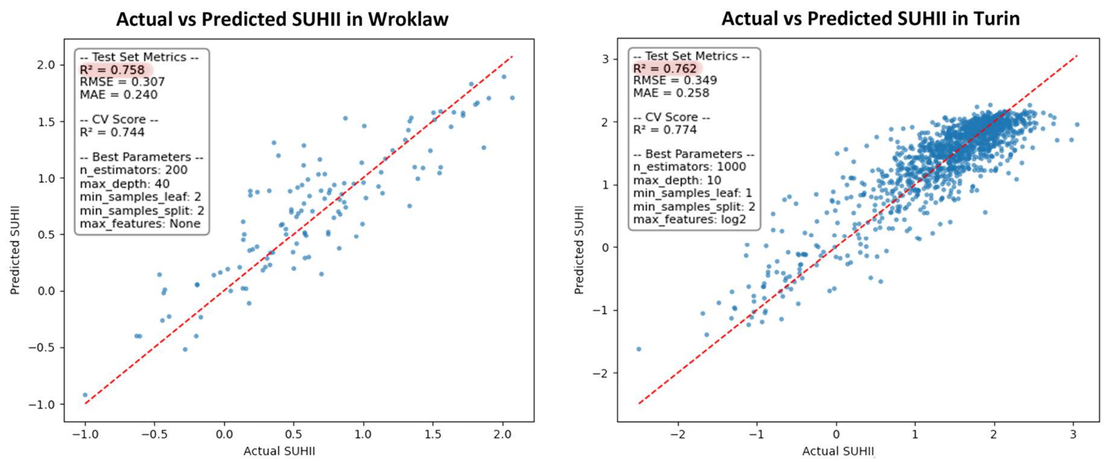


Figure 4.5. Observed versus predicted SUHII values for the independent test sets in Wrocław and Turin.

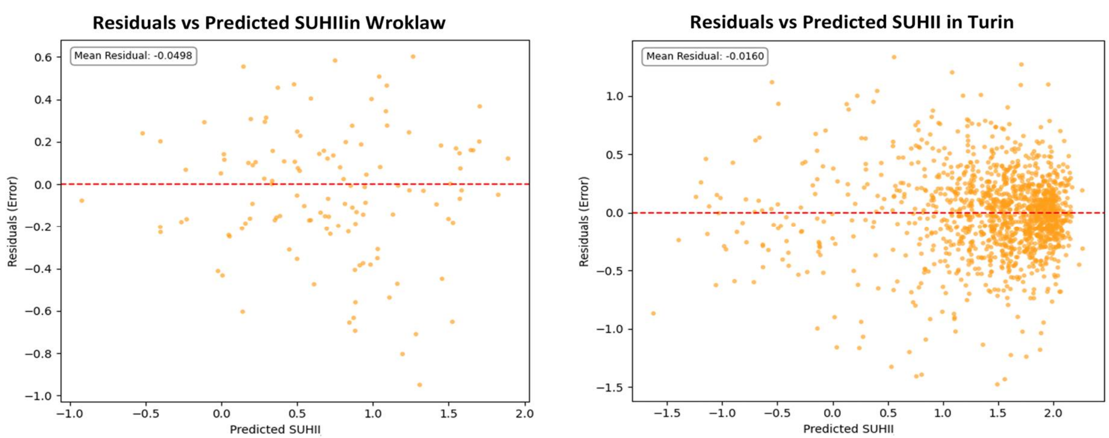


Figure 4.6. Residuals versus predicted SUHII values, used to inspect visible bias or uneven error patterns in the test-set predictions.

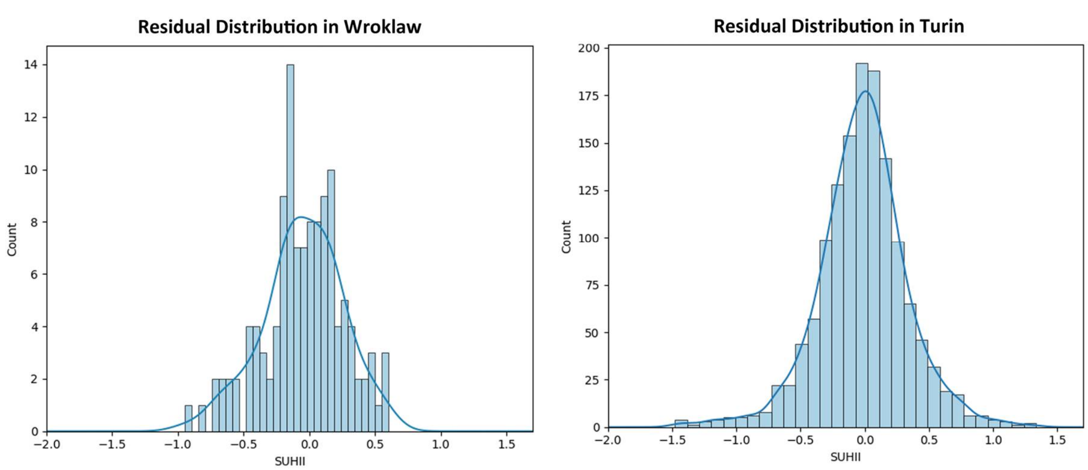


Figure 4.7. Residual distribution for the Wrocław and Turin Random Forest models.

The observed-versus-predicted plots showed a clear relation between measured and estimated SUHII values, consistent with the reported R^2 scores. The residual plots were used to check whether prediction errors were concentrated in specific predicted ranges. The residual histograms provided an additional view of the error distribution around zero. These diagnostics support the use of the calibrated models for interpreting warm-season SUHII patterns, while the final spatial prediction maps are discussed separately in Section 4.5.

4.4 Feature Importance: Identifying the Drivers of SUHII

Feature importance was extracted from the calibrated Random Forest models to compare the relative contribution of the selected predictors in Wrocław and Turin. The analysis was carried out only for the final warm-season models. It does not report seasonal repetitions of feature importance, because the other seasonal scenes were used for spatial assessment rather than for separate predictive models.

Both city models used the same ten predictors. Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio. In the Turin dataset, the BCR had been recalculated prior to the final modelling phase to ensure that it represented a valid ratio between building footprint area and census-section area. After this check, the Turin model kept BCR and S/V Ratio, so it had the same predictor structure as Wrocław.

The feature-importance values denote the size of each predictor in decreasing the prediction error in the Random Forest model. They should thus be interpreted as model-based indicators of relative predictive contribution. They do not on their own establish direct physical causality. Thus, the rankings are interpreted in conjunction with the results of the spatial assessment and the planning discussion in Chapter 5.

The comparative feature-importance ranking is reported in Table 4.6 and visualized in Figure 4.8.

Table 4.6. Comparative ranking of predictor variable importance for the warm-season Wrocław and Turin SUHII Random Forest models.

Rank	Wrocław variable	Wrocław importance	Turin variable	Turin importance
1	NDBI	0.3049	Emissivity	0.1759
2	NDMI	0.2359	PVI	0.1640
3	PVI	0.1029	NDWI	0.1489
4	Emissivity	0.0823	NDVI	0.1449
5	BCR	0.0698	NDBI	0.1408

6	NDVI	0.0551	NDMI	0.1323
7	NDWI	0.0473	Albedo	0.0317
8	Albedo	0.0405	BCR	0.0287
9	SVF	0.0315	SVF	0.0256
10	SV_Ratio	0.0299	SV_Ratio	0.0072

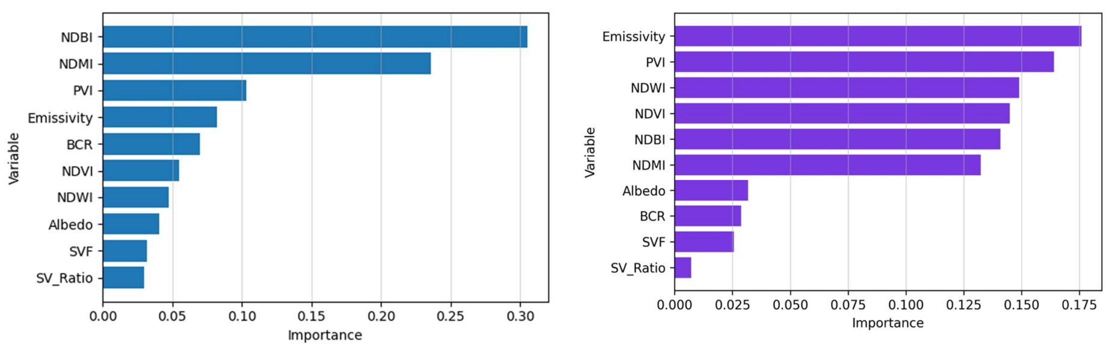


Figure 4.8. Feature-importance scores for the warm-season SUHII Random Forest models in Wrocław and Turin.

In Wrocław, NDBI had the highest importance value (0.3049), followed by NDMI (0.2359), PVI (0.1029), Emissivity (0.0823), and BCR (0.0698). This ranking suggests that, in the Wrocław warm-season model, built-up intensity and surface moisture conditions had the strongest predictive role, while vegetation proportion, emissivity, and building coverage also contributed to the estimation of SUHII.

In Turin, Emissivity had the highest importance value (0.1759), followed by PVI (0.1640), NDWI (0.1489), NDVI (0.1449), NDBI (0.1408), and NDMI (0.1323). This ranking indicates that radiative surface behavior, vegetation condition, water-related surface signal, built-up intensity, and moisture conditions were the main predictors in the Turin warm-season model.

The morphology variables were kept in both models, but their relative importance was lower than that of most of the spectral and radiative variables. The BCR in Wrocław was 5th, the SVF and S/V Ratio had lower values. BCR, SVF and S/V Ratio also appeared in the model but were ranked lower than the main spectral, vegetation, moisture and emissivity indicators in Turin.

Overall, the feature importance results show that the two cities have a similar set of relevant predictors, but in a different order of importance. In Wrocław the model is more strongly connected with moisture and built-up intensity. In Turin the ranking has a higher weight on the emissivity and vegetation/moisture indicators. In Chapter 5, these differences are used to support a planning interpretation of surface heat patterns in each city.

4.5 Spatial Prediction Maps of SUHII

After the Python-based training and validation, the model outputs were linked back to the GIS workflow to produce spatial prediction maps of SUHII. This step was carried out in ArcGIS Pro using the city-specific Random Forest results for the warm-season representative scenes of Wrocław and Turin.

This step was performed to visualize the predicted SUHII values over space. The validation metrics presented in Section 4.3 quantify the model performance, while the prediction maps illustrate the locations of higher and lower modeled SUHII values within each city. The maps thus assist in linking the model results to the spatial structure of the urban fabric.

Prediction layers were derived from the same active predictors as for Random Forests models: Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR and S/V Ratio. LST was not employed as an independent variable in the SUHII as the dependent variable. The maps produced are modeled SUHII for the warm season scenes and should be viewed in conjunction with the test set metrics and feature importance results presented in the previous sections.

Figure 4.9. SUHII spatial prediction for Wrocław and Turin. The maps are not substitutes for the original LST, SUHI, UTFVI, or hotspot layers in Chapter 3. Instead, they offer a model-based view on how the chosen predictors predict SUHII across the mapped spatial units.

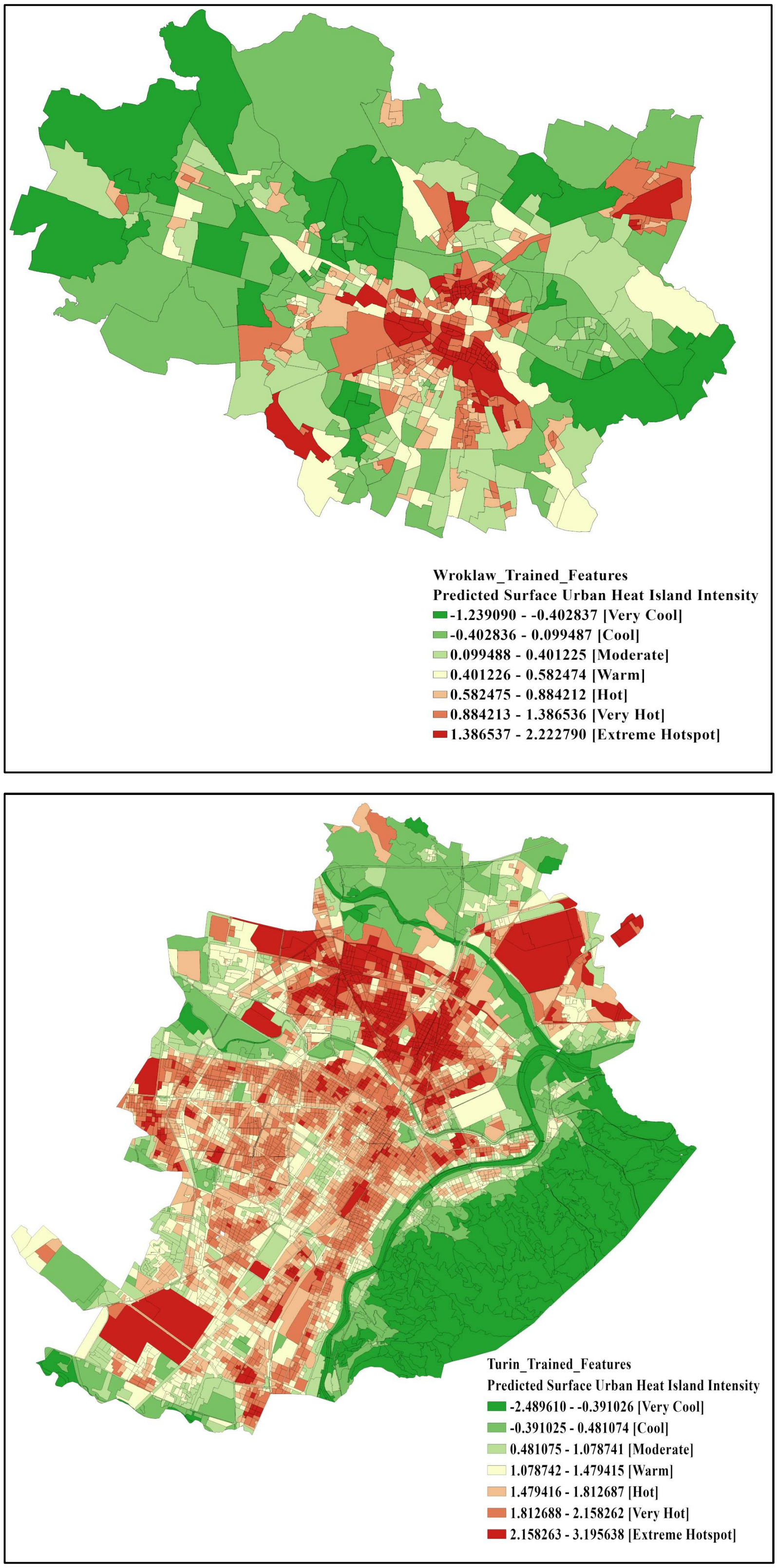


Figure 4.9. Spatial prediction of warm-season SUHII in Wrocław and Turin, derived from the calibrated Random Forest models.

Overall, this step connected the Python modeling results with the GIS-based spatial interpretation. Python was used for model calibration, validation, diagnostics, and feature-importance extraction, while ArcGIS Pro was used to visualize the predicted SUHII values as map outputs. These prediction maps support the planning-oriented discussion in Chapter 5, where they are read together with SUHI magnitude, UTFVI stress, hotspot recurrence, and feature-importance rankings.

5. Urban regeneration and SUHI

The previous chapters presented the two main analytical parts of the thesis. Chapter 3 described the multi-seasonal spatial assessment of LST, SUHII, SUHI, UTFVI, and thermal hotspots. Chapter 4 presented the warm-season Random Forest modeling of SUHII and the relative importance of the selected predictors. This chapter uses these results to discuss how surface thermal evidence can support planning interpretation in Wrocław and Turin.

The chapter does not introduce a new model or a new thermal index. Instead, it reads the spatial outputs together with the Random Forest results and the planning concepts reviewed in Chapter 1. SUHI is used here as the main planning-oriented thermal layer because it expresses the surface temperature difference between each urban location and the rural reference condition. UTFVI and hotspot recurrence are used as supporting layers to identify areas where relative thermal stress appears more clearly or more repeatedly across the selected seasonal scenes [91].

Outdoor thermal comfort is only considered as a planning consequence of surface overheating. This thesis does not calculate human thermal comfort indices such as PET, UTCI or PMV, nor does it use microclimatic simulation tools such as ENVI-met, SOLWEIG or RayMan. The analysis is based on surface thermal patterns derived from satellite data, GIS-based spatial assessment and Random Forest modelling of SUHII .

For this reason, the results should not be interpreted as direct measurements of pedestrian-level comfort. They identify areas of elevated surface thermal stress and help indicate where planning attention may be needed. These outputs can support later microclimate simulations or field measurements, but they do not replace them.

In this thesis, urban regeneration is considered from a thermal-planning perspective. The term refers to possible planning measures that may reduce surface overheating, such as strengthening blue-green infrastructure, increasing shading, improving surface permeability through de-sealing, introducing vegetation where feasible, and using cool or reflective materials in dense and sealed areas [92].

These measures are discussed in relation to the spatial patterns identified in the SUHI, UTFVI, hotspot, and SUHII outputs. The thesis does not measure the before-and-after effect of completed regeneration

[91] Kara et al., "Multi-Index Assessment of Surface Urban Heat Island (SUHI) Dynamics in Samsun Using Google Earth Engine"; Sobrino and Irakulis, "A Methodology for Comparing the Surface Urban Heat Island in Selected Urban Agglomerations Around the World from Sentinel-3 SLSTR Data."

[92] Maryam F. Abdulateef and Hoda A. S. Al-Alwan, "The Effectiveness of Urban Green Infrastructure in Reducing Surface Urban Heat Island," *Ain Shams Engineering Journal* 13, no. 1 (2022): 101526, https://doi.org/10.1016/j.asej.2021.06.012; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Sinha and Banerjee, "A Systematic Review of Land Use and Land Cover Changes and Their Influence on Urban Heat Island Dynamics."

projects. Instead, it uses the mapped thermal evidence and the model-based interpretation of SUHII drivers to suggest which types of planning measures appear more relevant for each case-study city.

5.1 SUHI magnitude and regeneration-oriented interpretation of thermal hotspots

The SUHI evaluation described in Chapter 3 provides the spatial basis for the planning interpretation developed in this section. In the thesis workflow, SUHII was used as the standardized thermal anomaly and as the target variable for the Random Forest models. SUHI was used differently. It was calculated as the difference between LST and the mean rural-reference LST, derived through the ESA WorldCover-based rural mask, and was used to interpret the urban–rural surface thermal contrast [93].

The SUHI, UTFVI, and hotspot outputs were already produced during the multi-seasonal spatial assessment. In this section, they are read together to identify where surface thermal stress appears more relevant from a planning perspective. This interpretation does not measure the effect of completed regeneration projects and does not provide a before-and-after assessment. It uses the selected Landsat scenes as representative thermal evidence for discussing possible planning responses.

The interpretation follows four steps, shown in Figure 5.1. First, SUHI mapping describes the magnitude of the rural–urban surface temperature difference. Second, UTFVI is used to classify relative ecological thermal stress from LST variation [94]. Third, seasonal hotspot maps identify areas where positive SUHI and high UTFVI stress occur in the same scene. Fourth, the seasonal hotspot layers are combined to identify areas where thermal stress appears more repeatedly across the selected dates.

[93] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Sobrino and Irakulis, "A Methodology for Comparing the Surface Urban Heat Island in Selected Urban Agglomerations Around the World from Sentinel-3 SLSTR Data."

[94] Kara et al., "Multi-Index Assessment of Surface Urban Heat Island (SUHI) Dynamics in Samsun Using Google Earth Engine."

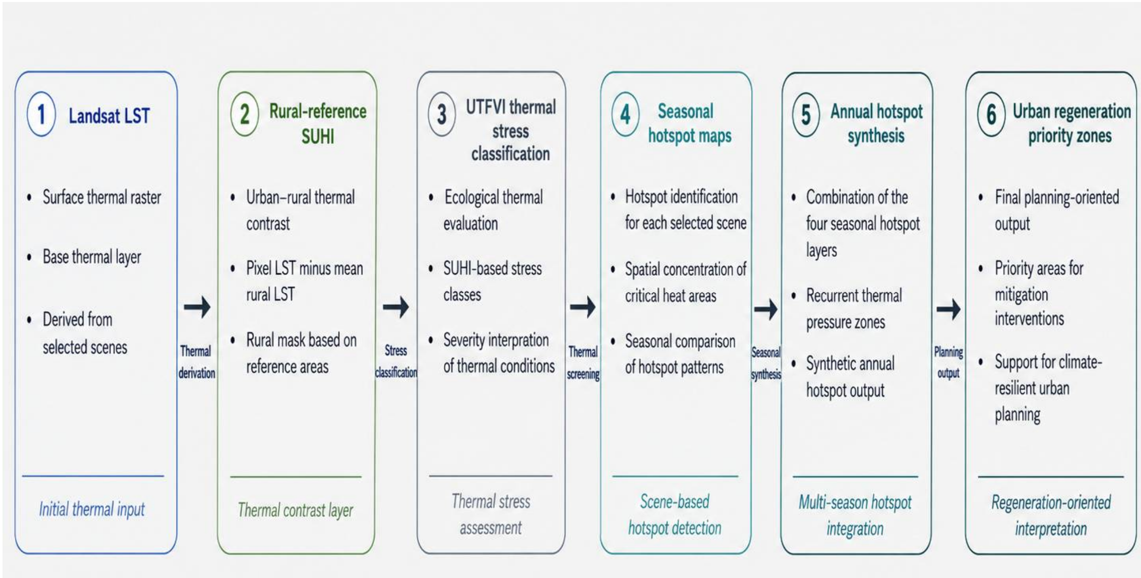


Figure 5.1. Planning interpretation workflow based on SUHI magnitude, UTFVI stress, seasonal hotspot identification, and annual hotspot synthesis.

In Wrocław, the interpretation of SUHI magnitude is linked to the relationship between built-up areas, open spaces, and blue-green infrastructure. Areas where positive SUHI overlaps with high UTFVI classes and recurrent hotspots can indicate locations where built-up surfaces reduce the cooling role of nearby open, green, or riparian areas [95]. From a planning perspective, these areas can be considered suitable for measures that strengthen blue-green continuity, increase tree canopy where possible, protect river-related cooling corridors, and improve surface permeability through de-sealing [96].

In Turin, the interpretation is more closely related to dense and sealed urban fabric. Where SUHI, UTFVI stress, and hotspot recurrence overlap, the spatial evidence suggests that artificial surfaces and limited evaporative capacity contribute to higher surface thermal stress [97]. Possible planning responses include targeted urban greening, green roofs where building conditions allow, cool or reflective materials, and permeability improvements in dense or post-industrial areas [98]. These measures are discussed as planning recommendations, not as interventions whose cooling effect was directly simulated in this thesis.

[95] Tasan et al., "Long-Term Dynamics of Urban Heat Island and Hot Spots in Wrocław."
[96] Abdulateef and A. S. Al-Alwan, "The Effectiveness of Urban Green Infrastructure in Reducing Surface Urban Heat Island."
[97] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[98] Mutani and Todeschi, "Roof-Integrated Green Technologies, Energy Saving and Outdoor Thermal Comfort"; Sinha and Banerjee, "A Systematic Review of Land Use and Land Cover Changes and Their Influence on Urban Heat Island Dynamics."

The comparison between Wrocław and Turin does not evaluate the success of previous regeneration projects. It compares how thermal stress is spatially distributed in two different urban contexts. In Wrocław, the planning interpretation points mainly to the protection and reinforcement of blue-green networks. In Turin, it points more directly to the need to improve dense built-up areas through surface-material changes, additional vegetation, and permeability where feasible [99].

This interpretation is also consistent with the Random Forest results from Chapter 4. In Wrocław, NDBI and NDMI had the highest importance values, indicating the role of built-up intensity and surface moisture in the warm-season SUHII model. In Turin, Emissivity, PVI, NDWI, NDVI, NDBI, and NDMI were the main predictors, suggesting a stronger role of radiative surface behavior together with vegetation and moisture-related variables. However, the planning interpretation in this chapter is not based on feature importance alone. It also considers the spatial overlap between SUHI magnitude, UTFVI stress, and recurrent hotspots.

Table 5.1 summarizes the main thermal evidence and planning implications for the two case studies.

Table 5.1. Summary of thermal evidence and regeneration-oriented planning implications for Wrocław and Turin.

Case Studies	Thermal evidence from spatial assessment	Regeneration-oriented interpretation	Planning implication
Wroclaw	Higher SUHI overlapping with UTFVI stress and recurrent hotspots	Built-up surfaces may locally reduce the cooling role of open spaces and blue-green areas	Strengthen blue-green continuity, increase vegetation and shading, protect riparian corridors, and promote de-sealing
Turin	Higher SUHI overlapping with UTFVI stress and recurrent hotspots	Dense and sealed urban fabric contributes to persistent surface thermal stress	Prioritize targeted greening, green roofs where feasible, cool or reflective materials, and permeability improvements

Overall, SUHI magnitude was used to identify areas where urban surfaces were warmer than the rural reference condition. UTFVI and hotspot recurrence helped distinguish isolated warm areas from locations where relative thermal stress appeared more repeatedly. Together, these layers support the identification of possible planning priority areas in Wrocław and Turin, while the actual performance of specific interventions would require further simulation or field-based assessment.

5.2 Comparative insights for SUHI-oriented urban planning

The comparison between Wrocław and Turin shows that SUHI-oriented planning should be adapted to the local urban and environmental context. Both cities were analyzed with the same methodological

[99] Esposito et al., "A Multi-City Statistical Modelling of Surface Urban Heat Island."

workflow, but the spatial patterns of surface heat and the model-based predictor rankings differed between the two case studies. For this reason, the planning interpretation should not be based on a single standard mitigation approach for both cities [100].

The results from SUHI, UTFVI, and hotspot mapping help identify where surface thermal stress appears more relevant spatially. The Random Forest results from Chapter 4 add another level of interpretation by showing which predictors had stronger importance in the warm-season SUHII models. These two types of evidence should be read together. Feature importance alone does not define planning priorities, and hotspot maps alone do not explain the physical drivers of SUHII. Their combined interpretation provides a more complete basis for discussing possible mitigation directions [101].

In both cities, the relevant variables belong to a similar group: built-up intensity, surface moisture, vegetation condition, and radiative surface properties. In the models, these conditions were represented mainly by NDBI, NDMI, NDVI, PVI, NDWI, Emissivity, and Albedo. However, their relative importance and planning meaning changed between Wrocław and Turin. This confirms that the same type of variable can have a different spatial role depending on the local urban structure, surface composition, and climatic setting [102].

In Wrocław, the planning interpretation is mainly related to the continuity of blue-green infrastructure. The spatial results suggest that thermal stress becomes more relevant where built-up areas interrupt or weaken the cooling role of open spaces, vegetation, and river-related corridors. Where recurrent hotspots overlap with higher SUHI magnitude and UTFVI stress, possible planning measures include reinforcing blue-green connections, increasing tree canopy where feasible, protecting riparian cooling corridors, and improving surface permeability through de-sealing [103].

In Turin, the interpretation is more strongly connected to dense and sealed urban surfaces. The feature-importance ranking gave a higher role to Emissivity, together with vegetation and moisture-related indicators such as PVI, NDWI, NDVI, and NDMI. This suggests that surface material behavior and limited evaporative capacity are important for understanding the warm-season SUHII pattern. In these areas, possible planning responses include targeted greening, green roofs where technically feasible,

[100] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[101] El Kihel et al., "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania"; Oukawa et al., "Fine-Scale Modeling of the Urban Heat Island."
[102] Sinha and Banerjee, "A Systematic Review of Land Use and Land Cover Changes and Their Influence on Urban Heat Island Dynamics."
[103] Abdulateef and A. S. Al-Alwan, "The Effectiveness of Urban Green Infrastructure in Reducing Surface Urban Heat Island"; Tasan et al., "Long-Term Dynamics of Urban Heat Island and Hot Spots in Wrocław."

cool or reflective materials, and permeability improvements, especially in dense or post-industrial sectors [104].

The comparison also shows that vegetation-based measures should not be interpreted in the same way in both cities. In Wrocław, vegetation and de-sealing are most relevant when they help maintain or reconnect existing cooling systems. In Turin, vegetation is more closely linked to increasing evaporative surfaces inside a dense and highly artificialized urban fabric. Similarly, material-based measures have a stronger role in Turin, where sealed surfaces and radiative properties appear more important in the model results, while in Wrocław they should be considered together with the protection of blue-green continuity [105].

A hotspot based interpretation may help to focus planning attention on those areas where is an overlap of several thermal indicators. Sites with positive SUHI, high UTFVI stress and repeated hotspot presence can be considered as more thermally critical than sites detected by only one indicator. However, these areas should be regarded as possible priority areas for further planning analysis and not as final intervention areas confirmed by simulation. More sophisticated design decisions would require some local data, field measurements or microclimatic modeling.

Overall, the comparative interpretation suggests two different planning directions. In Wrocław, the results support strategies focused on blue-green continuity, shading, and permeability. In Turin, they support strategies focused on surface-material improvement, targeted greening, green roofs where feasible, and permeability in dense urban areas. In both cases, the integration of SUHI, UTFVI, hotspot recurrence, and SUHII model interpretation provides a spatial basis for discussing urban heat mitigation while remaining within the methodological limits of the thesis.

[104] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics"; Yupeng Wang and Hashem Akbari, "The Effects of Street Tree Planting on Urban Heat Island Mitigation in Montreal," *Sustainable Cities and Society* 27 (November 2016): 122–28, https://doi.org/10.1016/j.scs.2016.04.013.

[105] Mutani and Todeschi, "Roof-Integrated Green Technologies, Energy Saving and Outdoor Thermal Comfort."

6. Conclusions: Integrated Reading and Interpretation of Results

6.1 Integrated interpretation across seasons, cities, and urban form

The analysis of Wrocław and Turin shows that surface thermal patterns are influenced by the combined effect of seasonal conditions, land cover, surface materials, moisture availability and urban form. Four representative Landsat scenes were selected for each city to represent summer, fall, winter and spring conditions. These scenes were used to analyze the spatial distribution of LST, SUHII, SUHI, UTFVI and recurrent hot spots.

The multi-seasonal assessment revealed the patterns of change in surface temperature during the year. The warm season scenes, however, showed the clearest surface thermal contrast and were used for the Random Forest modeling stage. This decision allowed the thesis to focus the predictive analysis to the period where differences in SUHII were more evident, while the other seasonal layers remained part of the spatial assessment and planning interpretation.

The final interpretation of the thesis is based on three connected levels. The first level is the multi-seasonal spatial assessment, which uses LST, SUHII, SUHI, UTFVI, and hotspot recurrence to describe where surface thermal stress appears and how it changes between selected seasons. This level provides the spatial evidence needed to identify areas where heat patterns are more recurrent.

The second level is the warm-season Random Forest modeling of SUHII. The models were developed separately for Wrocław and Turin and produced good test-set results, with R^2 values of 0.7576 for Wrocław and 0.7841 for Turin. These results indicate that the selected predictors explained a large part of the warm-season SUHII variation in both cities. The final predictor set included Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio. LST was excluded from the predictors because SUHII was derived from LST.

The third level is the planning interpretation. The spatial outputs and the Random Forest results were read together with the planning concepts discussed in Chapter 1 and the regeneration-oriented interpretation developed in Chapter 5. In this level, LCZ was used only as a rule-based interpretive reference for discussing planning priorities. It was not included as a Random Forest predictor.

This structure is important because the thesis does not rely on a single output to explain urban heat. The seasonal maps describe spatial and temporal differences. The Random Forest models identify which selected variables had stronger predictive importance in the warm-season SUHII models. The planning interpretation then connects these findings to possible mitigation measures, while remaining within the limits of the method.

Overall, the results show that surface thermal stress in both cities is related to the replacement of vegetated and permeable surfaces with artificial materials, the availability of moisture, the condition of vegetation, surface radiative behavior, and selected aspects of urban morphology. The relative importance of these factors differs between Wrocław and Turin, which means that planning responses should be adapted to the local urban and environmental context rather than applied uniformly.

6.1.1 Common surface drivers of SUHII Intensity of built-up, moisture, vegetation and radiative behaviour

The feature-importance results of the warm-season Random Forest models show that a similar group of surface-related variables influenced both case studies, but the order of importance was different for Wrocław and Turin. Variables are built up intensity, surface moisture, vegetation condition and radiative surface behaviour. Rankings should be viewed as model-based measures of relative predictive contribution, not as direct evidence of physical causality.

In Wrocław, the importance value of NDBI was highest (0.3049), then NDMI (0.2359), PVI (0.1029), Emissivity (0.0823) and BCR (0.0698). This suggests that the Wrocław warm season model was mostly associated with built-up intensity and moisture conditions. The contribution of PVI and Emissivity also shows that vegetation proportion and surface radiative behavior were relevant for estimating SUHII . BCR ranked fifth, which shows that the building coverage played an important role but less than the main spectral and moisture-related indicators.

The importance of Emissivity was the highest (0.1759) in Turin, followed by PVI (0.1640), NDWI (0.1489), NDVI (0.1449), NDBI (0.1408), and NDMI (0.1323). The ranking shows that the Turin model was more linked to surface radiative behavior and to indicators of vegetation or moisture. BCR (0.0287), SVF (0.0256) and S/V Ratio (0.0072) were included in the model, but their relative importance was lower than that of spectral and radiative variables.

The results indicate that the two cities share similar categories of relevant predictors, but not the same hierarchy. In Wrocław, the model gives more weight to the intensity of construction and the availability of moisture. In Turin the ranking is more weighted to emissivity, vegetation condition, water-related surface signal and moisture. This difference is consistent with the hypothesis that SUHII is controlled by similar physical processes in the two cities, but their relative importance depends on the local surface composition and urban structure.

Interpretation is sensitive in particular to moisture-related variables. NDMI was highly important in Wrocław and still relevant in Turin, while NDWI was one of the strongest predictors in Turin. These indicators suggest that the availability of moisture and the evaporative potential are closely connected with lower or higher surface thermal intensity. However, they should be interpreted in the context of local land cover and urban form, since the same index can represent different spatial conditions in different cities.

Vegetation indicators also have a context-dependent interpretation. Both PVI and NDVI were relevant in both models, but should not be interpreted as isolated cooling variables. They are affected by the continuity of vegetation, the presence of moisture, shading conditions and the surrounding built fabric. In Wroclaw vegetation is more related to the continuity of blue-green infrastructure. In Turin it is more a question of the need to increase evaporating and shaded surfaces in dense and sealed urban areas.

Overall, the feature-importance results confirm that built-up intensity, moisture, vegetation and radiative surface behavior are central in the interpretation of SUHII in both case studies. Simultaneously, the different rankings suggest that mitigation strategies should not be transferred from one city to another in a mechanical way. The model results should be read in parallel with the SUHI, UTFVI, hotspot and planning evidence presented in Chapter 3 and 5.

6.1.2 City-specific interpretation: blue-green connectivity and surface-material properties

Although Wrocław and Turin share similar categories of SUHII predictors, the relative importance of these variables differs between the two models. This means that the same methodological workflow produced comparable outputs, but the interpretation of the results must remain specific to each urban context.

In Wrocław, NDBI had the highest importance value (0.3049), while NDMI was the second most important predictor (0.2359). BCR also contributed to the model with an importance value of 0.0698. These results suggest that the Wrocław warm-season SUHII pattern was closely related to built-up intensity, surface moisture conditions, and, to a lesser extent, building coverage. From a planning perspective, this supports an interpretation focused on the relationship between built-up areas and the continuity of open, permeable, green, and river-related spaces .[106].

This reading is consistent with the spatial reading discussed in Chapter 5. In Wrocław, the commonalities of recurrent hot spots with higher SUHI and UTFVI stress could point to locations where the built-up surfaces break or reduce the cooling function of blue-green infrastructure. Planning measures should therefore focus on the protection of existing cooling corridors, better continuity between green and blue spaces, increase shading wherever possible and reduce unnecessary surface sealing.

The model came up with another ranking in Turin. Emissivity was the most important variable (0.1759), followed by PVI (0.1640), NDWI (0.1489), NDVI (0.1449), NDBI (0.1408), and NDMI (0.1323). This indicates that the SUHII pattern over Turin in the warm season was more strongly correlated with surface radiative behavior, vegetation condition, water-related surface signal, built-up intensity, and moisture. The Turin model was updated and validated with BCR and S/V Ratio but their relative importance values were lower than spectral and radiative indicators.

[106] Abdulateef and A. S. Al-Alwan, “The Effectiveness of Urban Green Infrastructure in Reducing Surface Urban Heat Island”; Tasan et al., “Long-Term Dynamics of Urban Heat Island and Hot Spots in Wrocław.”

For Turin, the planning interpretation therefore gives more attention to dense and sealed urban fabric, surface materials, and the limited availability of evaporative surfaces. Possible measures include cool or reflective materials, targeted greening, green roofs where feasible, and permeability improvements in dense or post-industrial areas [107]. These measures should be understood as planning directions derived from the spatial and model-based evidence, not as tested intervention scenarios.

When we compare the results we can see that the same set of predictors can suggest different planning priorities. In Wrocław, the results support a reading based on blue-green continuity, water retention and protection of open or riparian spaces. The results from Turin support an interpretation based on surface materials, vegetation and water availability and the improvement of sealed urban areas. These differences confirm that the strategies for SUHI mitigation should be designed for local urban structure and surface conditions, rather than directly transposed from one city to another.

Finally, the city-specific interpretation links the Random Forest findings with the multi-seasonal spatial evidence. Feature importance illustrates the variables that had more predictive contribution in the warm-season models while SUHI, UTFVI, and hotspot recurrence show where the surface thermal stress was spatialized. Reading outputs together provides a more reliable basis for discussing planning priorities in Wroclaw and Turin.

6.2 Planning implications and future directions for SUHI research

The Landsat–GEE–QGIS–Python–ArcGIS Pro workflow developed in this thesis provides a reproducible structure for analyzing SUHI patterns at the urban scale. The workflow combines multi-seasonal spatial assessment with warm-season Random Forest modeling, allowing surface thermal patterns to be interpreted together with land-cover, vegetation, moisture, radiative, and morphology-related variables [108].

The planning value of the workflow is mainly related to the combination of different thermal outputs. SUHI magnitude describes the surface temperature difference from the rural reference condition. UTFVI can be used to interpret relative ecological thermal stress. Hotspot recurrence shows where thermal stress is more recurrent among the selected seasonal scenes. These layers, together with the feature-importance results from Chapter 4, can be read to provide a useful basis to discuss possible planning priorities in Wrocław and Turin (Kara et al., 2025).

Results for Wrocław show that in planning, it is necessary to consider maintenance and enhancement of blue-green continuity, improvement of permeability, and protection of open or riparian spaces contributing to surface cooling. The significance of NDBI and NDMI in the warm-season model further

[107] Esposito et al., "A Multi-City Statistical Modelling of Surface Urban Heat Island"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[108] Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."

confirms this interpretation, signifying the significance of built-up intensity and moisture availability to the SUHII pattern.

For Turin, the results indicate a stronger role of surface radiative behavior, vegetation condition, water-related surface signal, built-up intensity, and moisture. Emissivity was the first-ranked predictor, followed by PVI, NDWI, NDVI, NDBI and NDMI . This helps planning measures concerning cool or reflective materials, targeted greening, green roofs where possible, permeability improvements in dense or sealed urban areas [109].

These implications for planning should be considered within the boundaries of the thesis. The analysis does not model cooling effect of specific interventions, and does not assess completed regeneration projects. It provides surface thermal patterns and model-based predictor relationships that can be used to inform further planning assessment. Detailed design of interventions would require further local data, field measurements, microclimatic simulations and feasibility analysis.

Several future research directions could extend the work developed in this thesis:

1. Seasonal predictive modeling. Future research could apply Random Forest modeling to autumn, winter, and spring scenes as well as summer scenes. This would help examine whether predictor importance changes across the year and whether different variables become more relevant under different seasonal conditions.
2. Additional three-dimensional morphology variables. The model could be improved by including a broader set of urban-form indicators, such as building height, height-to-width ratio, street canyon geometry, volumetric density, and façade exposure. These variables could help describe the three-dimensional structure of the urban fabric in more detail [110].
3. In-situ validation. Future work could compare satellite-derived surface thermal patterns with air-temperature measurements, mobile transects, or fixed meteorological stations. This would help clarify the relationship between SUHI, canopy-layer thermal conditions, and pedestrian-level heat exposure [111].
4. Human activity and anthropogenic heat proxies. Comparable data on population density, traffic intensity, energy use, or nighttime lights could be added if available for both cities at a

[109] Esposito et al., "A Multi-City Statistical Modelling of Surface Urban Heat Island"; Mutani et al., "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics."
[110] Yicong Chen et al., "The Impacts and Thresholds Detection of 2D/3D Urban Morphology on the Heat Island Effects at the Functional Zone in Megacity during Heatwave Event," *Sustainable Cities and Society* 118 (January 2025): 106002, https://doi.org/10.1016/j.scs.2024.106002; Xu et al., "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity."
[111] Oukawa et al., "Fine-Scale Modeling of the Urban Heat Island."

consistent spatial scale. This would make it possible to examine the role of human activity without reducing the comparability between Wrocław and Turin.

5. Direct evaluation of regeneration effects. Future studies could use reliable historical datasets or monitored intervention areas to compare pre- and post-regeneration thermal conditions. This would allow the cooling performance of specific measures, such as de-sealing, urban greening, green roofs, or cool materials, to be assessed more directly.
6. Outdoor thermal comfort modeling. A later stage of research could connect the SUHI results with microclimatic simulation tools or thermal comfort indices such as PET, UTCI, or PMV. This would help move from surface thermal assessment toward pedestrian-level outdoor comfort analysis, which was outside the methodological scope of this thesis.

Overall, this thesis shows that combining satellite remote sensing, GIS processing, and Random Forest regression can support the interpretation of SUHI patterns in different urban contexts. The comparison between Wrocław and Turin confirms that similar variable groups can be relevant in both cities, but their relative importance and planning meaning depend on local surface conditions, urban form, and climatic setting. The main contribution of the work is therefore methodological and interpretive: it provides a structured way to map surface thermal stress, model warm-season SUHII, and discuss planning responses without claiming direct simulation of regeneration effects.

7. Bibliography

Abdulateef, Maryam F., and Hoda A. S. Al-Alwan. "The Effectiveness of Urban Green Infrastructure in Reducing Surface Urban Heat Island." *Ain Shams Engineering Journal* 13, no. 1 (2022): 101526. https://doi.org/10.1016/j.asej.2021.06.012.

Chanpichaigosol, N., C. Chaichana, and D. Rinchumphu. "Urban Heat Island Classification through Alternative Normalized Difference Vegetation Index." *Global Journal of Environmental Science and Management* (Tehran, Iran) 11, no. 1 (2025): 57–76. https://doi.org/10.22034/gjesm.2025.01.04.

Chen, Yicong, Weibo Ma, Yamei Shao, et al. "The Impacts and Thresholds Detection of 2D/3D Urban Morphology on the Heat Island Effects at the Functional Zone in Megacity during Heatwave Event." *Sustainable Cities and Society* 118 (January 2025): 106002. https://doi.org/10.1016/j.scs.2024.106002.

El Kihel, Manar, Tudor Caciora, Cristian Grava, Radu-Cătălin Țarcă, Laila Saafadi, and Abdessalam Ouallali. "Integrating Multi-Source Remote Sensing and Ensemble Machine Learning to Quantify the Surface Urban Heat Island in Oradea Municipality, Romania." *Urban Climate* 67 (June 2026): 102872. https://doi.org/10.1016/j.uclim.2026.102872.

Esposito, Antonio, Gianluca Pappaccogli, Fabio Bozzeda, and Riccardo Buccolieri. "A Multi-City Statistical Modelling of Surface Urban Heat Island: Application to Italian Cities." *Urban Climate* 64 (December 2025): 102717. https://doi.org/10.1016/j.uclim.2025.102717.

Esposito, Antonio, Gianluca Pappaccogli, Antonio Donateo, et al. "Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves: A Study of Milan and Lecce (Italy)." *Remote Sensing* 16, no. 23 (2024): 4496. https://doi.org/10.3390/rs16234496.

Haashemi, Sirous, Qihao Weng, Ali Darvishi, and Seyed Kazem Alavipanah. "Seasonal Variations of the Surface Urban Heat Island in a Semi-Arid City." *Remote Sensing* 8, no. 4 (2016): 352. https://doi.org/10.3390/rs8040352.

Hidalgo García, David, and Julián Arco Díaz. "Modeling of the Urban Heat Island on Local Climatic Zones of a City Using Sentinel 3 Images: Urban Determining Factors." *Urban Climate* 37 (May 2021): 100840. https://doi.org/10.1016/j.uclim.2021.100840.

Howard, Luke. *THE CLIMATE OF LONDON*. n.d.

Kara, Yiğitalp, Veli Yavuz, and Anthony R. Lupo. "Multi-Index Assessment of Surface Urban Heat Island (SUHI) Dynamics in Samsun Using Google Earth Engine." *Atmosphere* 16, no. 6 (2025): 712. https://doi.org/10.3390/atmos16060712.

Li, Huidong, Yuyu Zhou, Xiaoma Li, et al. "A New Method to Quantify Surface Urban Heat Island Intensity." *Science of The Total Environment* 624 (May 2018): 262–72. https://doi.org/10.1016/j.scitotenv.2017.11.360.

Morabito, Marco, Alfonso Crisci, Giulia Guerri, Alessandro Messeri, Luca Congedo, and Michele Munafò. "Surface Urban Heat Islands in Italian Metropolitan Cities: Tree Cover and Impervious Surface Influences." *Science of The Total Environment* 751 (January 2021): 142334. https://doi.org/10.1016/j.scitotenv.2020.142334.

Mutani, Guglielmina, Alessandro Scalise, Xhoana Sufa, and Stefania Grasso. "Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics: A Comprehensive Modelling Approach." *Atmosphere* 15, no. 12 (2024): 1435. https://doi.org/10.3390/atmos15121435.

Mutani, Guglielmina, and Valeria Todeschi. "Roof-Integrated Green Technologies, Energy Saving and Outdoor Thermal Comfort: Insights from a Case Study in Urban Environment." *International Journal of Sustainable Development and Planning* 16, no. 1 (2021): 13–23. https://doi.org/10.18280/ijsdp.160102.

Mutani, Guglielmina, Valeria Todeschi, and Simone Beltramino. "Improving Outdoor Thermal Comfort in Built Environment Assessing the Impact of Urban Form and Vegetation." *International Journal of Heat and Technology* 40, no. 1 (2022): 23–31. https://doi.org/10.18280/ijht.400104.

Oke, T. R. "The Energetic Basis of the Urban Heat Island." *Quarterly Journal of the Royal Meteorological Society* 108, no. 455 (1982): 1–24. https://doi.org/10.1002/qj.49710845502.

Oukawa, Gabriel Yoshikazu, Patricia Krecl, and Admir Créso Targino. "Fine-Scale Modeling of the Urban Heat Island: A Comparison of Multiple Linear Regression and Random Forest Approaches." *Science of The Total Environment* 815 (April 2022): 152836. https://doi.org/10.1016/j.scitotenv.2021.152836.

Peng, Shushi, Shilong Piao, Philippe Ciais, et al. "Surface Urban Heat Island Across 419 Global Big Cities." *Environmental Science & Technology* 46, no. 2 (2012): 696–703. https://doi.org/10.1021/es2030438.

Rasul, Azad, Heiko Balzter, Claire Smith, et al. "A Review on Remote Sensing of Urban Heat and Cool Islands." *Land* 6, no. 2 (2017): 38. https://doi.org/10.3390/land6020038.

Sinha, Sudarshana, and Ankhi Banerjee. "A Systematic Review of Land Use and Land Cover Changes and Their Influence on Urban Heat Island Dynamics." *Discover Geoscience* 3, no. 1 (2025): 253. https://doi.org/10.1007/s44288-025-00359-4.

Sobrino, José Antonio, and Itziar Irakulis. "A Methodology for Comparing the Surface Urban Heat Island in Selected Urban Agglomerations Around the World from Sentinel-3 SLSTR Data." *Remote Sensing* 12, no. 12 (2020): 2052. https://doi.org/10.3390/rs12122052.

Stewart, I. D., and T. R. Oke. "Local Climate Zones for Urban Temperature Studies." Bulletin of the American Meteorological Society. *Bulletin of the American Meteorological Society* 93, no. 12 (2012): 1879–900. https://doi.org/10.1175/BAMS-D-11-00019.1.

Tasan, Melika, Jolanta Dąbrowska, Krystyna Michałowska, and Anna Uciechowska-Grakowicz. "Long-Term Dynamics of Urban Heat Island and Hot Spots in Wrocław: A 25-Year Satellite-Based Analysis Using Machine Learning." *Sustainable Cities and Society* 132 (September 2025): 106797. https://doi.org/10.1016/j.scs.2025.106797.

Wang, Yupeng, and Hashem Akbari. "The Effects of Street Tree Planting on Urban Heat Island Mitigation in Montreal." *Sustainable Cities and Society* 27 (November 2016): 122–28. https://doi.org/10.1016/j.scs.2016.04.013.

Wen, Cong, Ali Mamtimin, Haitao Wang, et al. "Nonlinear Drivers of Canopy and Surface Urban Heat Islands across China's Climate Zones: Marginal and Interaction Effects from Explainable Machine Learning." *Sustainable Cities and Society* 142 (May 2026): 107307. https://doi.org/10.1016/j.scs.2026.107307.

Xi, Yantao, Shuangqiao Wang, Yunxia Zou, XingChi Zhou, and Yuanfan Zhang. "Seasonal Surface Urban Heat Island Analysis Based on Local Climate Zones." *Ecological Indicators* 159 (February 2024): 111669. https://doi.org/10.1016/j.ecolind.2024.111669.

Xu, Duo, Yiquan Wang, Dian Zhou, Yupeng Wang, Qian Zhang, and Yujun Yang. "Influences of Urban Spatial Factors on Surface Urban Heat Island Effect and Its Spatial Heterogeneity: A Case Study of Xi'an." *Building and Environment* 248 (January 2024): 111072. https://doi.org/10.1016/j.buildenv.2023.111072.

Zhou, Decheng, Jingfeng Xiao, Stefania Bonafoni, et al. "Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives." *Remote Sensing* 11, no. 1 (2018): 48. https://doi.org/10.3390/rs11010048.

8. Sitography

Copernicus Climate Data Store. ERA5-Land hourly data from 1950 to present. Used for the extraction of acquisition-hour meteorological variables, including 2 m air temperature, 2 m dewpoint temperature, 10 m wind components, and surface solar radiation downwards.

Available at: Climate Data Store

ESA WorldCover 2021 v2 . European Space Agency . Used for land cover classification and to define the rural reference mask in the calculation of SUHI. The rural classes retained were Tree cover, Shrubland, Grassland, Cropland, Bare/sparse vegetation, Herbaceous wetland and the rural reference did not include Built-up areas and Permanent water bodies.

Available at: ESA World Cover

Google Earth Engine Data Catalog Designed for access and processing of Landsat 8/9 Collection 2 Level 2 imagery and ESA WorldCover 2021 v200 data within the Google Earth Engine environment.

Available at: Earth Engine Data Catalog

Google Earth Engine Platform The package is used for image filtering, pre-processing, spectral index calculation, LST and SUHII derivation, SUHI raster generation and export in GeoTIFF.

Available at: Google Earth Engine

Landsat 8/9 C2 L2 United States Geological Survey/NASA Primary satellite data source for surface reflectance indices, emissivity, LST, SUHII, and SUHI-related raster layers.

Available at: usgs.gov

Meteostat. used to support the comparative climatic context of Wroclaw and Turin using historical climate and weather data.

Available at: Meteostat

PVGIS Photovoltaic Geographical Information System. . Joint Research Centre of the European Commission. Used as an external reference to check the plausibility of solar irradiance values.

Available at: PHOTOVOLTAIC GEOGRAPHICAL INFORMATION SYSTEM

APPENDIX

The appendix contains supporting technical and spatial material related to the workflow developed in this thesis. It includes the main Python scripts used for meteorological data processing and machine-learning preparation, the hyperparameter-tuning code and outputs used to calibrate the Random Forest regression models, and supplementary spring and autumn thermal maps for the two case-study cities.

These materials are provided to document the reproducibility of the analytical workflow and to clarify the computational and spatial outputs that support the results presented in Chapters 3 and 4. The appendix does not introduce new methodological assumptions beyond those discussed in the main text. Instead, it provides technical documentation and selected supplementary maps that help verify how the main analytical outputs were produced.

A. Python scripts

This appendix contains an overview of the main Python scripts used during the data-processing and modeling stages of the thesis. The full executable scripts are provided as supplementary digital files, together with the final GIS and modeling outputs. Such structure was used to preserve thesis appendix readability and to avoid formatting issues related to long code blocks in Word. The scripts supported four parts of the main work flow. The first script processed ERA5-Land NetCDF files to extract meteorological variables for each selected Landsat acquisition date. The second script prepared the census based modeling datasets exported out of QGIS. Model diagnostics and feature importance plots were generated using the third script. The fourth script did Random Forest hyper-parameter tuning.

The code descriptions below document the main computational logic used in the thesis. The supplementary ***.py*** files should be used as the reference version for execution.

A.1 ERA5-Land meteorological data processing script

Complete supplementary file: A1_ERA5_Land_processing.py

The ERA5-Land processing script was used to prepare the meteorological values reported in Table 3.4. For each city and selected Landsat date, the script opened the corresponding NetCDF file and extracted the variables needed to describe the acquisition-hour atmospheric conditions.

The main processing steps were:

1. Open the ERA5-Land NetCDF file for each city and selected date.
2. Extract 2 m air temperature, 2 m dewpoint temperature, 10 m u and v wind components, and surface solar radiation downwards.

3. Convert 2 m air temperature from Kelvin to degrees Celsius.
4. Calculate wind speed from the 10 m u and v wind components.
5. Calculate relative humidity from air temperature and dewpoint temperature using the Magnus approximation.
6. Convert accumulated surface solar radiation downwards into mean hourly solar irradiance by differencing consecutive hourly accumulation values and dividing by 3600 seconds.
7. Select the hourly interval closest to the Landsat acquisition period, corresponding to 10:00–11:00 UTC.
8. Calculate the spatial mean of each variable for the selected interval.
9. Export the final table.

This script produced the acquisition-hour meteorological table used in Section 3.2.3. The final thesis table reports air temperature, wind speed, mean hourly solar irradiance, and relative humidity for the eight selected Landsat scenes.

A.2 Dataset cleaning and Random Forest preparation script

Complete supplementary file: A2_RF_data_preparation.py

The dataset preparation script was used to clean the census-based attribute tables exported from QGIS and prepare them for Random Forest regression. The same workflow was applied separately to Wrocław and Turin.

The main processing steps were:

1. Load the final census-based CSV file exported from QGIS.
2. Standardize variable names so that the same terminology was used in Python, QGIS, and the thesis text.
3. Select the warm-season variables used for predictive modeling.
4. Define the final predictor set: Albedo, Emissivity, NDBI, NDMI, NDVI, NDWI, PVI, SVF, BCR, and S/V Ratio.
5. Define SUHII as the target variable.
6. Exclude LST from the predictor set because SUHII was derived from LST.
7. Remove incomplete records after selecting the final predictors and target.
8. Split the dataset into training and testing subsets using an 80/20 ratio.

9. Train and evaluate the Random Forest regression model.
10. Export feature-importance values and diagnostic prediction tables.

This script provided the structured input used for the modeling results reported in Chapter 4.

A.3 Model evaluation and diagnostic plot script

Complete supplementary file: A3_model_diagnostics.py

The model diagnostic script was used to visualize the results of the Random Forest regression models for Wrocław and Turin. It used the diagnostic prediction tables and feature-importance outputs generated during the modeling stage.

The main outputs were:

1. Observed versus predicted SUHII plots.
2. Residuals versus predicted SUHII plots.
3. Residual distribution plots.
4. Feature-importance diagrams.

These outputs were used to support the model evaluation and interpretation presented in Sections 4.3 and 4.4.

B. Hyperparameter tuning (code and results)

Complete supplementary file: B_RF_hyperparameter_tuning.py

The Random Forest hyperparameter-tuning procedure was implemented separately for Wrocław and Turin. Separate tuning was used because the two datasets differed in sample size, spatial-unit structure, and predictor distributions.

The tuning process tested different combinations of the following Random Forest parameters:

1. Number of trees.
2. Maximum tree depth.
3. Number of predictors considered at each split.
4. Minimum number of samples required to split a node.
5. Minimum number of samples required at a terminal leaf node.

RandomizedSearchCV was used with five-fold cross-validation. The selected hyperparameters and cross-validation results are reported in Tables 4.1 and 4.2 of the main thesis.

The Wrocław model used the following selected parameters: 200 trees, maximum depth of 40, all predictors considered at each split, minimum samples split of 2, and minimum samples leaf of 2. The best cross-validation R^2 was 0.7441.

The Turin model used the following selected parameters: 1000 trees, maximum depth of 30, log2 predictors considered at each split, minimum samples split of 15, and minimum samples leaf of 1. The best cross-validation R^2 was 0.7737.

```
B.1 Wrocław tuning output

Best parameters:
n_estimators = 200
max_depth = 40
max_features = None
min_samples_split = 2
min_samples_leaf = 2

Best cross-validation R² = 0.7441
```

```
B.2 Turin tuning output

Best parameters:
n_estimators = 1000
max_depth = 30
max_features = log2
min_samples_split = 15
min_samples_leaf = 1

Best cross-validation R² = 0.7737
```

C. Supplementary seasonal maps

In this appendix section, we present additional thermal raster outputs for spring and autumn from the multi-seasonal spatial assessment. The main text includes the main representative maps applied for the explanation of the workflow and to support the planning-oriented interpretation, while the supplementary maps document the transitional-season thermal anomaly and ecological stress layers for both Wrocław and Turin.

The maps presented in this appendix are mainly focused on SUHII and UTFVI. SUHII is the standardized thermal anomaly in each study area and is the main modeling target, while UTFVI helps to interpret the ecological stress based on relative LST variation. These layers were retained as they offer the most direct link between the seasonal thermal assessment and the analytical framework presented in Chapters 3 and 4.

This appendix does not repeat LST, SUHI, land-cover indices and seasonal hotspot maps. The LST was already described in the methodology as the base thermal layer from which the SUHII and UTFVI were derived. The rural-reference workflow described in the main text provides the SUHI, and the annual composite hotspot map continues to be the primary hotspot output for identifying persistent priority areas for planning and mitigation. Land-cover indices are not reiterated here since illustrative examples are already given in the main text. The appendix deals with transitional-season thermal anomaly and ecological stress layers.

C.1 Wrocław supplementary spring and autumn maps

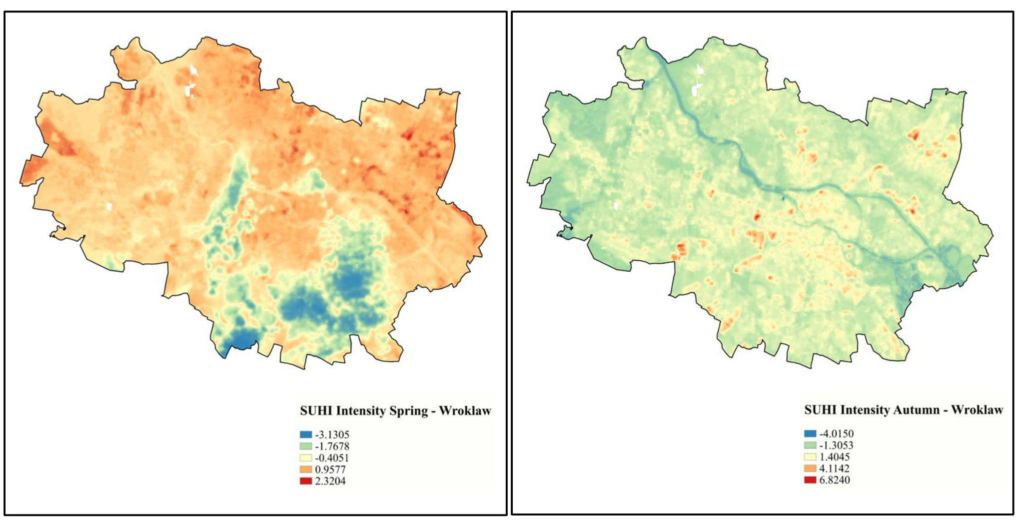


Figure C.1. Supplementary spring and autumn Surface Urban Heat Island Intensity (SUHII) maps for Wrocław, showing standardized thermal anomalies relative to the mean LST of each Landsat scene.

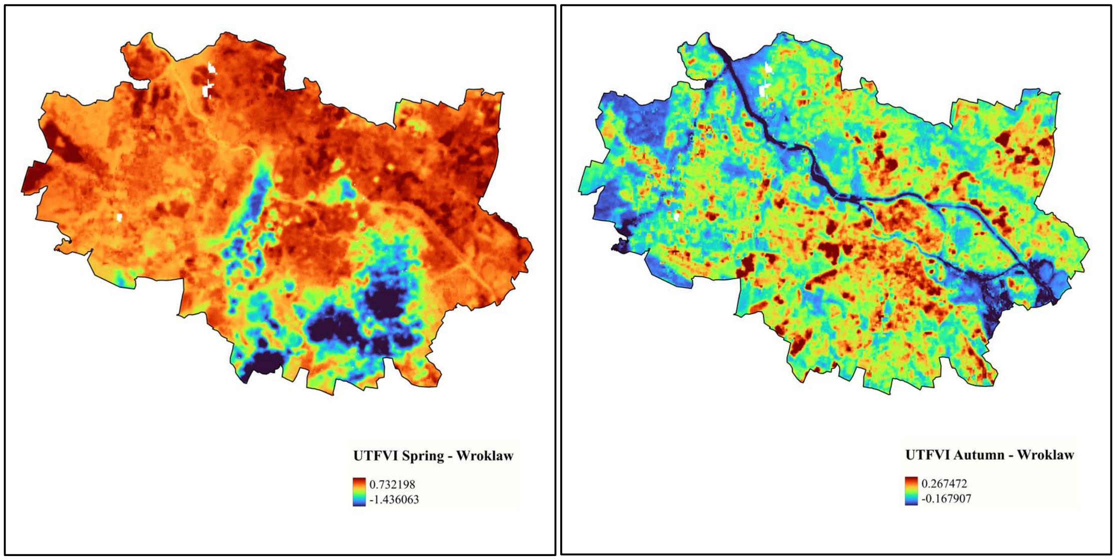


Figure C.2. Supplementary spring and autumn Urban Thermal Field Variance Index (UTFVI) maps for Wrocław, showing ecological thermal stress patterns derived from the relative LST variation of each Landsat scene.

C.2 Turin supplementary spring and autumn maps

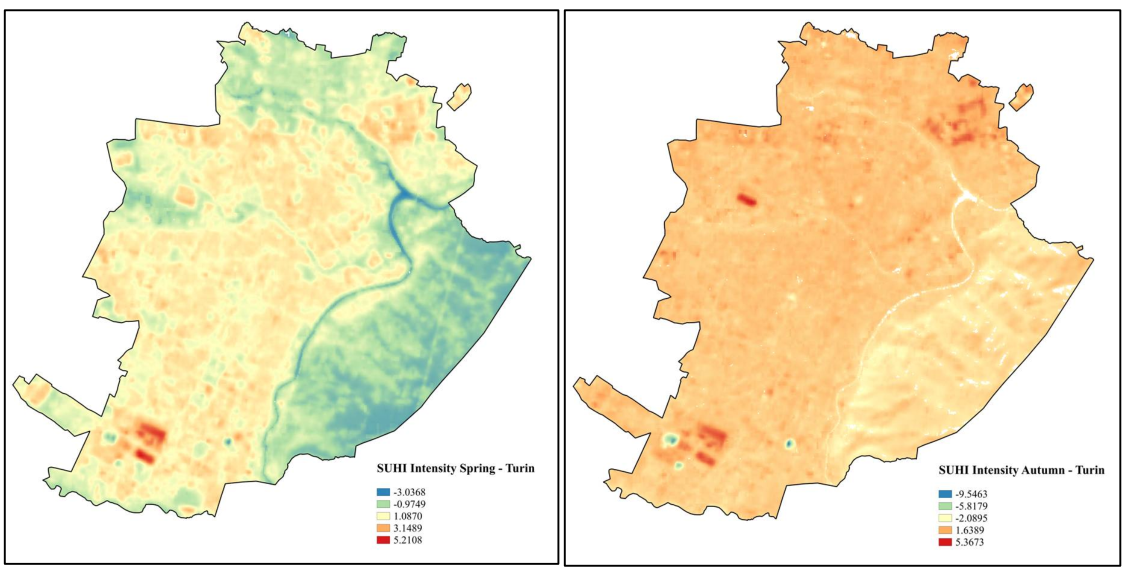
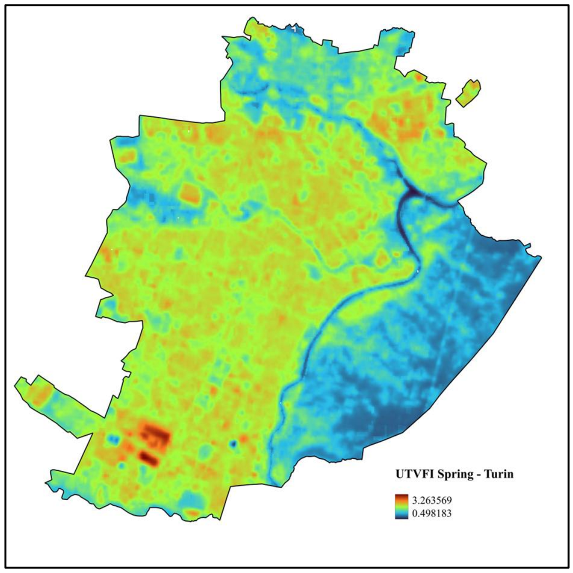


Figure C.3. Supplementary spring and autumn Surface Urban Heat Island Intensity (SUHII) maps for Turin, showing standardized thermal anomalies relative to the mean LST of each Landsat scene.



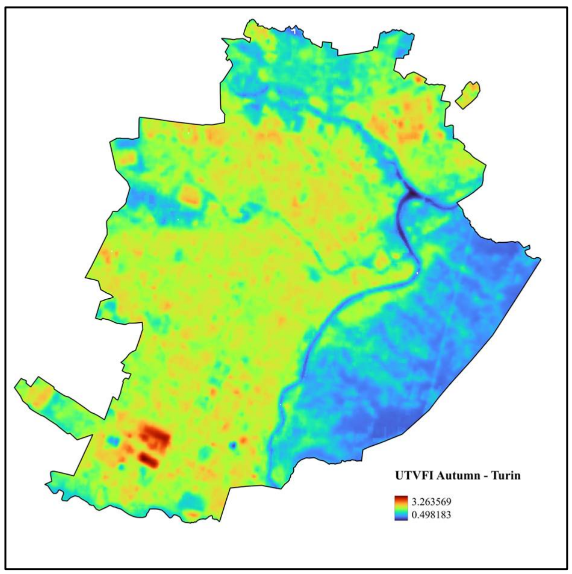


Figure C.4. Supplementary spring and autumn Urban Thermal Field Variance Index (UTFVI) maps for Turin, showing ecological thermal stress patterns derived from the relative LST variation of each Landsat scene.

D. Literature Table

No	Title	Author(s)	Publication Year	Place / Localization	Scale	Object	Factors	Tools / Analysis	Results (as cited in text)
1	The effectiveness of urban green infrastructure in reducing surface urban heat island	Abdulateef, M. F., & A. S. Al-Alwan, H.	2022	Not specified	City / Local	Urban Green Infrastructure (UGI) & SUHI	BGI continuity, canopy cover	Not specified	De-sealing and canopy expansion can support blue-green infrastructure continuity and contribute to SUHI mitigation.
2	Urban heat island classification through alternative normalized difference vegetation index	Chanpichaigosol, N., Chaichana, C., & Rinchumphu, D.	2025	Not specified	Not specified	UHI classification	NDVI, NDWI	Remote Sensing	Higher vegetation density is generally associated with lower LST through shading and evapotranspiration.
3	Integrating multi-source remote sensing and ensemble machine learning to quantify the surface urban heat island...	El Kihel, M., Caciora, T., Grava, C., Țarcă, R.-C., Saafadi, L., & Ouallali, A.	2026	Oradea municipality, Romania	City	SUHI quantification	Land cover, biophysical variables, NDBI	Multi-source remote sensing, Ensemble ML	Machine-learning and remote-sensing methods can estimate and map SUHII patterns; built-up intensity is identified as an important explanatory factor.
4	A multi-city statistical modelling of surface urban heat island: Application to Italian cities	Esposito, A., Pappaccogli, G., Bozzeda, F., & Buccolieri, R.	2025	Italian cities	Multi-city / Regional	SUHI modeling	Albedo, reflective materials	Statistical modeling	Reflective or cool materials can reduce solar heat absorption and contribute to lower surface temperatures.
5	Urban Morphology and Surface Urban Heat Island Relationship During Heat Waves: A Study of Milan and Lecce (Italy)	Esposito, A., Pappaccogli, G., Donateo, A., Salizzoni, P., Maffeis, G., Semeraro, T., et al.	2024	Milan and Lecce, Italy	City	Urban Morphology & SUHI	SVF, BCR, Building Height	Remote Sensing	Three-dimensional urban form affects heat storage and longwave radiation exchange in dense areas.
6	Modeling of the Urban Heat Island on local climatic zones of a city using Sentinel 3 images: Urban determining factors	Hidalgo García, D., & Arco Díaz, J.	2021	Not specified	City	UHI on LCZs	Sky View Factor (SVF)	Sentinel 3	Low SVF can limit longwave heat loss after sunset and influence heat retention within the urban fabric.
7	THE CLIMATE OF LONDON	Howard, L.	1883 (n.d. in text)	London	City	Climate / Thermal anomalies	Not specified	Not specified	Provides an early reference for the study of urban thermal differences.
8	Multi-Index Assessment of Surface Urban Heat Island (SUHI) Dynamics in Samsun Using Google Earth Engine	Kara, Y., Yavuz, V., & Lupo, A. R.	2025	Samsun, Turkey	City	SUHI Dynamics	UTFVI	Google Earth Engine (GEE)	UTFVI can be used to identify areas with higher ecological and environmental thermal stress.
9	A new method to quantify surface urban heat island intensity	Li, H., Zhou, Y., Li, X., Meng, L., Wang, X., Wu, S., & Sodoudi, S.	2018	Not specified	Not specified	SUHII quantification	Impervious Surface Areas (ISA)	Linear regression	Linear regression between LST and impervious surface area has been used in conventional SUHI analysis.
10	Surface urban heat islands in Italian metropolitan cities: Tree cover and impervious surface influences	Morabito, M., Crisci, A., Guerri, G., Messeri, A., Congedo, L., & Munafò, M.	2021	Italian metropolitan cities	Multi-city / Regional	SUHI influences	Tree cover, impervious surfaces	Not specified	Tree cover and impervious surfaces are related to SUHI intensity in Italian metropolitan areas.

11	Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics: A Comprehensive Modelling Approach	Mutani, G., Scalise, A., Sufa, X., & Grasso, S.	2024	Turin	Not specified	UHI Dynamics	LST, LULC, SVF, Emissivity	Machine Learning, Remote Sensing	Combining machine learning, GIS, and remote sensing can support the spatial interpretation of SUHII patterns.
12	Roof-Integrated Green Technologies, Energy Saving and Outdoor Thermal Comfort: Insights from a Case Study...	Mutani, G., & Todeschi, V.	2021	Not specified	Local / Building	Outdoor Thermal Comfort	Green roofs, reflective materials	Not specified	Green roofs and cool materials can influence surface albedo and outdoor thermal conditions.
13	Improving Outdoor Thermal Comfort in Built Environment Assessing the Impact of Urban Form and Vegetation	Mutani, G., Todeschi, V., & Beltramino, S.	2022	Not specified	Local	Outdoor Thermal Comfort	Urban form, vegetation	Not specified	Urban form and vegetation are relevant factors for interpreting outdoor thermal conditions.
14	Urban Heat Island Mitigation: A GIS-based Model for Hiroshima	Mutani, G., Todeschi, V., & Matsuo, K.	2019	Hiroshima, Japan	City	UHI Mitigation	Building Coverage Ratio (BCR)	GIS	Building coverage and sealed surfaces are relevant for understanding urban heat and ventilation conditions.
15	The energetic basis of the urban heat island	Oke, T. R.	1982	Not specified	Not specified	UHI energetic basis	Surface energy balance	Not specified	Urban heat island formation is linked to changes in the surface energy balance.
16	Surface Urban Heat Island Across 419 Global Big Cities	Peng, S., Piao, S., Ciais, P., Friedlingstein, P., Ottle, C., Bréon, F.-M., et al.	2012	419 Global Big Cities	Global	SUHI	Vegetation, Anthropogenic heat	Not specified	Vegetation contributes to surface cooling through evapotranspiration, while human heat release can add to urban warming.
17	A Review on Remote Sensing of Urban Heat and Cool Islands	Rasul, A., Balzter, H., Smith, C., Remedios, J., Adamu, B., Sobrino, J. A., et al.	2017	Not specified	Not specified	Remote Sensing of UHI	Not specified	Review	Provides background for remote-sensing-based UHI and SUHI assessment.
18	A systematic review of land use and land cover changes and their influence on urban heat island dynamics	Sinha, S., & Banerjee, A.	2025	Not specified	Not specified	LULC changes and UHI	LULC	Systematic review	Land-use and land-cover changes help explain spatial differences in urban thermal intensity.
19	A Methodology for Comparing the Surface Urban Heat Island in Selected Urban Agglomerations Around the World...	Sobrino, J. A., & Irakulis, I.	2020	Selected Urban Agglomerations Worldwide	Global / Multi-city	SUHI comparison	LST (Rural/Urban contrast)	Sentinel-3 SLSTR	Provides formulas and methodological guidance for comparing urban LST with rural reference zones.
20	Local Climate Zones for Urban Temperature Studies	Stewart, I. D., & Oke, T. R.	2012	Not specified	Local / City	Local Climate Zones (LCZ)	Physical, radiative, metabolic factors	Not specified	LCZ classification supports the interpretation of urban thermal patterns in relation to built form and land cover.
21	Long-term dynamics of urban heat island and hot spots in Wrocław: A 25-year satellite-based analysis using machine learning	Tasan, M., Dąbrowska, J., Michałowska, K., & Uciechowska-Grakowicz, A.	2025	Wrocław, Poland	City	UHI and Hot spots	Blue-green infrastructure	Satellite analysis, ML	In Wrocław, built-up areas and blue-green infrastructure are relevant for interpreting hotspot distribution.

22	Evaluation of Urban-Scale Building Energy-Use Models and Tools—Application for the City of Fribourg, Switzerland	Todeschi, V., Boghetti, R., Kämpf, J. H., & Mutani, G.	2021	Fribourg, Switzerland	City	Building Energy-Use	BCR, S/V ratio	Urban-scale models	Building coverage and S/V ratio are useful morphology indicators in urban-scale modeling.
23	The effects of street tree planting on Urban Heat Island mitigation in Montreal	Wang, Y., & Akbari, H.	2016	Montreal, Canada	City	UHI mitigation	Street tree planting	Not specified	Street-tree planting can contribute to urban heat mitigation in artificialized urban areas.
24	Nonlinear drivers of canopy and surface urban heat islands across China's climate zones: Marginal and interaction effects...	Wen, C., Mamtimin, A., Wang, H., Wang, Y., Liu, H., Sayit, H., et al.	2026	China (various climate zones)	National	Drivers of canopy & SUHI	Population density, Nighttime lights	Explainable ML	Population density, nighttime lights, and other human-activity proxies can help explain canopy and surface heat patterns where comparable data are available.
25	Seasonal surface urban heat island analysis based on local climate zones	Xi, Y., Wang, S., Zou, Y., Zhou, X., & Zhang, Y.	2024	Not specified	Seasonal	Seasonal SUHI	LCZ, building heights	Not specified	LCZs and building-height characteristics are useful for interpreting seasonal SUHI differences.
26	Influences of urban spatial factors on surface urban heat island effect and its spatial heterogeneity: A case study of Xi'an	Xu, D., Wang, Y., Zhou, D., Wang, Y., Zhang, Q., & Yang, Y.	2024	Xi'an, China	City	Urban spatial factors & SUHI	3D urban form, SVF, BCR	Not specified	Three-dimensional urban geometry is associated with spatial variation in SUHI patterns.
27	Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives	Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., et al.	2019	Not specified	Not specified	SUHI Progress	Pervious/impervious surfaces	Satellite Remote Sensing	LST and air temperature represent different parts of the urban thermal environment, and land-cover change influences the surface energy balance.