Politecnico di Torino

Master's Degree program in
ARCHITECTURE FOR SUSTAINABILITY
A.a. 2025/2026
September 2026

Mapping, Understanding and Addressing Urban Heat

An Integrated Remote Sensing and Machine Learning Framework for Seasonal Surface Urban Heat Island Analysis, Driver Assessment and Targeted Mitigation in Turin.

Supervisor:
Guglielmina Mutani

Candidate:
Chaitali Amrutrao Kalokhe

Acknowledgements

I would like to start by thanking my supervisor, Prof. Guglielmina Mutani, who has been so kind and patient throughout the entire process of this thesis. Her expertise, guidance and feedback have helped in developing the direction and rigour of this work.

And my deepest gratitude to my dear parents, Mr. Amrutrao Kalokhe and Mrs. Sujata Kalokhe who have always supported and believed in me. And my sister, Mayuri, my inspiration in life and academically. They are the reason I could fulfil this dream of studying abroad.

To all my friends in Italy and back in my home, India, who have been my support system in this Master's journey which has been a great learning experience. I Thank you all.

Abstract

Climate change and urban expansion together intensify urban heat islands, a phenomenon well documented at large cities. City wide assessments that provide actionable mitigation based on its drivers are comparatively rare. This thesis addresses that gap by examining the Surface Urban Heat Island (SUHI) in Turin, Italy by combining the satellite remote sensing methods, global Local Climate Zone (LCZ) classification, locally derived urban morphology, and machine learning. The result helps in identifying the main drivers of the SUHI and SUHII and links them to spatially targeted mitigation and compensation measures.

Using ARPA Piemonte weather station records from 2021 to 2025, the study identifies a climatologically representative year. Landsat 8/9 Collection 2 imagery for four seasonal dates within that year is processed in Google Earth Engine to derive Land Surface Temperature (LST), a z-score based SUHI Intensity (SUHII), and a rural reference SUHI along with other predictor like NDVI, NDBI, NDMI, NDWI, PVI, albedo and surface emissivity. Following to this, the urban structure of Turin is characterised through the Demuzere et al. (2022) global LCZ classification at 100m resolution. The spectral, morphological and climate predictors of SUHI and SUHII for each season and for the year as a whole were ranked following Pearson correlation, variance inflation factor screening, and a Random Forest model interpreted through SHAP.

Across all seasons, it is found that top three consistent predictors are Normalised Difference Vegetation Index (NDVI), Impervious surface fraction (ISF) and albedo. Spectral predictors account for the highest contribution in summer whereas morphology assumes comparatively greater importance in spring

season. Climate variables show negligible importance for per season model because there is no spatial variation within a single season. However, in all seasons combined model, the climate variables are assigned a substantial importance. The SHAP dependence curves identify season specific turning points. NDVI ranges from 0.405 in spring to 0.574 in autumn, ISF from 0.504 to 0.607, and Albedo from 0.126 to 0.270. In summer, when the NDVI turning point is reached it results in a mean cooling effect of -1.87°C per cell, with ISF contributing further -0.81°C per cell. In winter, where albedo gains comparative importance, the per cell benefits are shifted to −0.34 °C for NDVI, −0.33 °C for Albedo, and −0.16 °C for ISF. Structurally constrained cells in the compact built core carry a mean annual residual of −1.11 °C per cell. This is addressed through the C.I.R.C.A. compensation catalogue of the Metropolitan City of Turin.

Season-specific driver attribution connects to a spatially explicit planning framework. giving authorities a reproducible, data-driven basis for prioritising where to direct the heat-reduction investment across Turin.

Keywords: Surface Urban Heat Island; Random Forest; SHAP; Landsat; Turin.

Table of Contents

Acknowledgements ..... **iii**

Abstract ..... **v**

List of Abbreviations ..... **x**

Chapter 1- Literature Review: Urban Heat Islands, Remote Sensing, Local Climate Zones and Machine Learning Approaches **2**

1.1 Urban Heat Islands: definitions and the canopy-layer vs. surface UHI 2

1.2 Remote sensing of land surface temperature and SUHI/SUHII intensity indices 5

1.3 The Local Climate Zone framework and global LCZ mapping (Demuzere et al., 2022) 7

1.4 Machine learning approaches to attributing morphological and climatic drivers of SUHI 9

1.5 Consequences of Surface Urban Heat Islands 11

1.6 Nature-based solutions and environmental compensation frameworks for UHI mitigation 13

1.7 Research gap and objectives of this thesis 15

Chapter 2- Turin as a Case Study: Urban Context, Monitoring Network and Research Workflow **18**

2.1 Geographical and climatic context of Turin 18

2.2 Ground-based monitoring network: ARPA regional stations and the Via della Consolata urban station 19

2.3 Overview of data sources: satellite imagery, the Demuzere LCZ dataset and Geoportale Torino 20

2.4 Methodological workflow: from typical-year selection to mitigation strategies 21

Chapter 3- Selecting a Representative Year and Deriving Remote-Sensing and Climate Indicators **24**

3.1 Selection of a climatologically typical year 24

3.2 Satellite scene selection and pre-processing 28

3.2.1 Overview of the Google Earth Engine processing workflow 28

3.3 Land Surface Temperature retrieval 32

3.4 Ancillary spectral and biophysical indicators 36

3.5 Surface Urban Heat Island intensity: the SUHII z-score and the rural-reference SUHI 42

3.6 Urban Thermal Field Variance Index 46

Chapter 4- Local Climate Zones and Urban Morphology of Turin **51**

4.1 The Demuzere et al. (2022) global LCZ classification and its adaptation to Turin 51
4.2 Locally derived morphological indicators 53
4.3 Relationship between the global LCZ classification and locally derived morphology 58
4.4 The rural reference station within the LCZ scheme 60

Chapter 5- Modelling the Drivers of Surface Urban Heat Island Intensity 61
5.1 Purpose and Analytical Approach 61
5.2 Modelling Dataset and Spatial Framework 62
5.3 Predictor Variables 64
5.4 Multicollinearity and VIF Analysis 66
5.5 Random Forest Modelling 72
5.6 Model Performance and Validation 72
5.7 Feature Importance and SHAP Interpretation 76
5.7.1 Summer 2025 77
5.7.2 Winter 2025 80
5.7.3 Spring 2025 82
5.7.4 Autumn 2025 85
5.7.5 All Seasons Combined 87
5.8 Seasonal and Cross-Season Interpretation 94
5.9 Predicted SUHII and SUHI 95
5.10 Implications for the Next Stage 98

Chapter 6 — Mitigation Guidance and Environmental Compensation 100
6.1 From Modelled Drivers to Intervention Design 100
6.2 SHAP Feature Importance and Seasonal Driver Profiles 101
6.2.1 NDVI Dependence Curve 103
6.2.2 Impervious Surface Fraction Dependence Curve 104
6.2.3 Albedo Dependence Curve and the Land-Cover Confound 105
6.3 Spatial Priority Classification: Count of Warming Features 107
6.4 LCZ-Based Feasibility Screening 110
6.5 Mitigation Potential on Feasible Cells 110
6.6 Residual Thermal Burden and the Compensation Requirement 112
6.7 Environmental Compensation and the C.I.R.C.A. Framework 114
6.8 Zone-Specific Recommendations 115
6.9 Spatial Priority Mapping 120

Chapter 7 — Discussion and Conclusions 123
7.1 Summary of Findings 123
7.2 Methodological Contributions 124
7.3 Limitations 125
7.4 Planning Guidance 126
7.5 Future Work 127
7.5.1 SHAP-Based Evaluation of Compensatory Measures 127

7.5.2 Expanded Feasibility Classification 128
7.5.3 Resolving the Albedo Confound 128
7.5.4 Multi-Year Temporal Generalisation 129
7.6 Concluding Remarks 129

References 131

Appendix 134
Appendix A - Google Earth Engine 134
Appendix B - Predictor Screening, Random Forest Modelling, and Mitigation Scenario Scripts 134
B.1 Predictor screening and full 22-predictor model – 134
B.2 Final 9-predictor seasonal models – 135
B.3 All Seasons Combined model – 135
B.4 SHAP based Mitigation modelling 135
Appendix C – Additional Maps and Per- Season Statistics Tables 135
C.1 SHAP-Based Mitigation Analysis — Additional Seasons 136
Winter 2025 136
Spring 2025 138
Autumn 2025 139

List of Abbreviations

The following abbreviations are used throughout this thesis.

ARPA	Agenzia Regionale per la Protezione Ambientale
BCR	Building Coverage Ratio
C.I.R.C.A.	Catalogo di Interventi di Rigenerazione e Compensazione Ambientale (Catalogue of Regeneration and Environmental Compensation Interventions)
DSM / DTM	Digital Surface Model / Digital Terrain Model
EIA / SEA	Environmental Impact Assessment / Strategic Environmental Assessment
GCR	Green Coverage Ratio
GEE	Google Earth Engine
GIS	Geographic Information System
H/W	Height-to-Width ratio (street canyon aspect ratio)
ISF	Impervious Surface Fraction
LCZ	Local Climate Zone
LSE	Land Surface Emissivity
LST	Land Surface Temperature
NBS	Nature-Based Solutions
NDBI	Normalised Difference Built-up Index
NDMI	Normalised Difference Moisture Index
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
PSF	Pervious Surface Fraction
PTGM	Piano Territoriale Generale Metropolitano (General Territorial Metropolitan Plan)
PVI	Proportion of Vegetation Index
RF	Random Forest
RMSE	Root Mean Square Error

SHAP	SHapley Additive exPlanations
SUHI	Surface Urban Heat Island
SUHII	Surface Urban Heat Island Intensity
SVF	Sky View Factor
UHI	Urban Heat Island
UHII	Urban Heat Island Intensity
UTFVI	Urban Thermal Field Variance Index
VIF	Variance Inflation Factor
WUDAPT	World Urban Database and Access Portal Tools

Chapter 1- Literature Review: Urban Heat Islands, Remote Sensing, Local Climate Zones and Machine Learning Approaches

1.1 Urban Heat Islands: definitions and the canopy-layer vs. surface UHI

Extreme heat has become one of the most pressing climate risks for European cities with each city facing a disproportionate share of burden. Analysis was done on some of the Europe's largest cities which projected that under a 3°C global warming scenario, the number of days reaching temperatures 35°C or greater can rise by 53%. Resulting in prolonged and more severe heat waves across these cities[1]. Across the European cities annual deaths due to heat is about 175,000 deaths. In Milan, Barcelona and ten other European cities about 2300 deaths were caused by a 10-day heat wave in June and July 2025[2]. Given their geographic and climatic proximity to Turin, Northern Italy's urban centres closely depict this pattern. Public-health costs are matched by economic ones. European cities lost about 0.3% and 0.5% of GDP annually with southern Europe exceeding to 1% and projects future losses up to five times today's level over the next three decades[3].

Air conditioning reaches only around one-fifth of European buildings at present[4]. This is where passive, nature-based approaches come in — vegetation, reflective surfaces, and climate-responsive urban form, the focus of this thesis — which cut the heat absorbed directly by buildings and urban surfaces, and rank

[1] **Invalid source specified.**
[2] **Invalid source specified.**
[3] **Invalid source specified.**
[4] **Invalid source specified.**

among the strategies current European policy discussion has put forward for managing the continent's growing cooling demand[5]

Urban heat island effect (UHI) is amongst the most documented effects of human land transformation on the local climate. The progressive substitution of vegetated, permeable surfaces with built fabric and sealed surfaces has reshaped the thermal balance of urban areas, sustaining temperatures that diverge measurably from those of the surrounding areas.[6] Recognition of the phenomenon dates back to the early nineteenth century, and since, it has generated a substantial body of research spanning meteorology, remote sensing, and urban planning.[7]

Confusion keeps cropping up in the literature because of the term "heat island" used to describe this phenomenon that actually belong to different atmospheric layers, and to the urban surface. The canopy-layer UHI captures the extra amount of air temperature found in the layer that extends from the ground up to the mean building height. Since this layer is closest to where people experience the heat, it is most relevant to thermal comfort. This canopy layer UHI is typically studies using the fixed and mobile weather stations. The boundary-layer UHI tracks the changes in air temperature above the roof level which is the effect shaped by the urban surfaces beneath it. This type of UHI is studied through towers, aircraft. Or radiosondes. The surface UHI (SUHI) refers to the differences in land surface temperatures rather than air temperature. To measure SUHI, thermal-infrared satellite sensors can be used at city-wide scale, though what it measures is different physical quantity than air temperature.[8] This

[5] **Invalid source specified.**

[6] (Oke, 1982)

[7] (Arnfield, 2003)

[8] (Voogt & Oke, 2003)

thesis works with the surface type of urban heat island effect and uses the term SUHI consistently to keep it separate from the canopy-layer air temperature studies.

The physical mechanisms behind this excess warmth are well established by now. The decrease in sky view factor within street canyons results in trapping the outgoing long-wave radiation at night. Surface materials like asphalt and concrete stores daytime heat and releases it at night due to its high thermal admittance. Impervious surface reduces evapotranspirative cooling compared to the vegetated surfaces or bare-soil surfaces. And built areas, traffic, and industry all add to anthropogenic heat as an enhances direct energy input.[9] These dynamics play out at the level of individual streets and blocks which is the exact spatial scale that the Local Climate Zone framework (outlined in Section 1.3) was built to map and evaluate.

[9] (Oke, 1982)

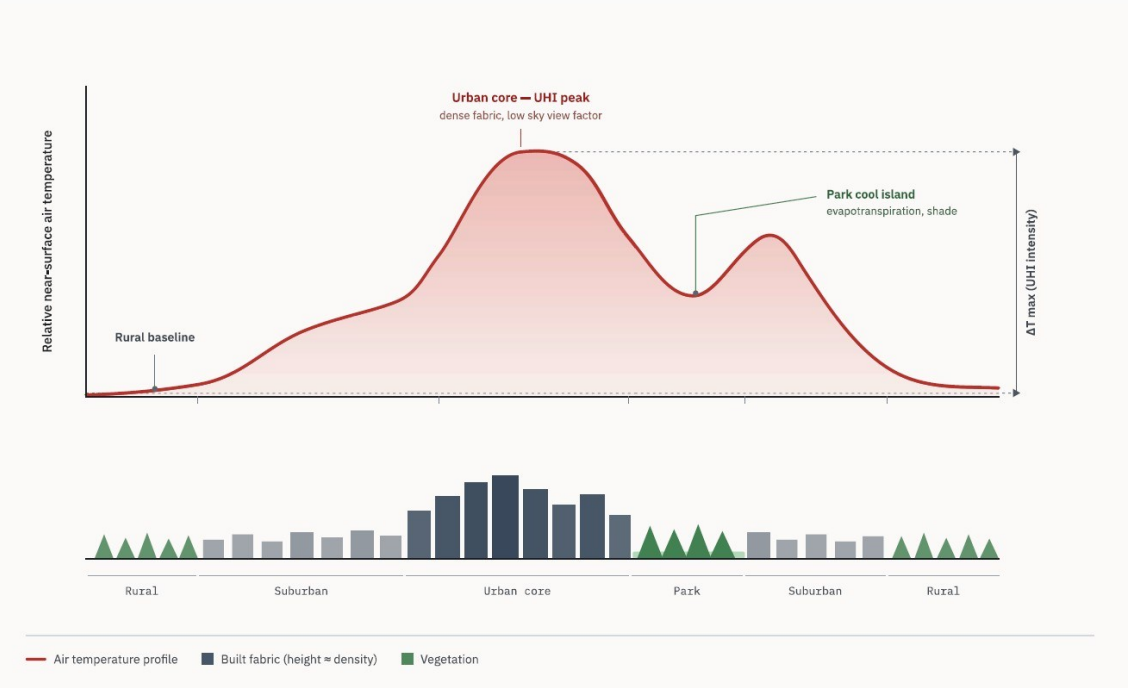


Figure 1.1. Schematic urban–rural temperature cross-section illustrating the canopy-layer UHI concept (idealised profile, not measured data).

1.2 Remote sensing of land surface temperature and SUHI/SUHII intensity indices

Thermal-infrared satellite sensors, including the Landsat Thematic Mapper/ETM+/TIRS series, MODIS, and Sentinel-3 SLSTR have made city-wide, temporally repeatable SUHI monitoring possible at spatial resolutions ranging from 10 metres to about a kilometre. Spatial detail and revisit frequency pull against each other across sensors: Landsat 8/9 delivers 100 m native thermal resolution, resampled to 30 m in the standard product, on a 16-day repeat cycle, while MODIS covers the same ground daily but only at roughly 1 km, too coarse to pick out individual Local Climate Zones. Converting top-of-atmosphere thermal radiance into land surface temperature calls for correcting atmospheric transmittance and surface emissivity, a task several single-channel and split-window algorithms have been built to handle and subsequently validated

against Landsat Collection 2 products.[10] Surface broadband albedo, used in this thesis as an ancillary indicator, is estimated from multispectral reflectance following the narrowband-to-broadband conversion approach of Liang.[11][12]

Once LST has been retrieved, the surface UHI intensity (SUHII) can be expressed in more than one way. The first, and most widely used in large multi-city comparisons, standardises each pixel's LST as a z-score relative to the spatial mean and standard deviation of the scene, producing a dimensionless index independent of the absolute temperature of a given day.[13] The second expresses SUHI in absolute degrees Celsius as the difference between an urban pixel's LST and that of a designated rural reference location. This approach is closer to the original ground-based UHI definition and more directly interpretable by non-specialist audiences, including the planning bodies this thesis's mitigation chapter ultimately addresses. Automated processing pipelines for deriving both LST and SUHI/SUHII from Landsat scenes in a reproducible way, the approach adopted for the Google Earth Engine workflow described in Chapter 3 have been demonstrated at city scale by Avdan and Jovanovska.[14]

Because the two SUHII formulations answer different questions: one asks *how anomalous is this pixel relative to the rest of the city today*, the other asks *how much warmer is the city than a specific rural benchmark.* Several researchers have argued that reporting only one of the two risks over- or under-stating the

[10] (Ermida, Soares, Mantas, Göttsche, & Trigo, 2020); (Wei & Sobrino, 2024)

[11] (Liang, 2001)

[12] Landsat 8/9 overpass Turin at approximately 10:00–10:30 local solar time, with a 16-day repeat cycle per satellite (partially offset when both satellites' acquisitions are combined). All LST, SUHII and SUHI values in this thesis therefore describe mid-morning conditions only; they are not directly comparable to afternoon-peak or night-time canopy-layer UHI values reported in some of the studies cited above. This constraint is revisited as a methodological limitation in Section 7.2.

[13] (Peng, et al., 2012); (Zhang, Li, & Liu, 2022)

[14] (Avdan & Jovanovska, 2016)

practical severity of the phenomenon.[15] This thesis therefore computes both the z-score SUHII and the absolute, rural-reference SUHI in parallel (Section 3.3), treating the two as complementary outputs from the outset. At the metropolitan scale, prior work in the Turin context by Mutani and Todeschi[16] has already shown that morphological and land-cover indicators derived at fine spatial resolution can explain a substantial share of intra-urban SUHI variability, providing a direct methodological precedent for the approach developed in Chapters 4–5.

Large-sample studies using global thermal datasets have further shown that SUHI intensity and its drivers vary systematically with background climate, vegetation phenology and city morphology.[17] This is one of the reasons this thesis derives indicators separately for four seasonal dates rather than a single summer snapshot.

1.3 The Local Climate Zone framework and global LCZ mapping (Demuzere et al., 2022)

The Local Climate Zone (LCZ) scheme was proposed by Stewart and Oke[18] as a standardised, ten-plus-seven-class typology of urban and natural land cover, defined by measurable surface-structure properties — building/tree height, plan-area density, and pervious/impervious fraction — independent of land use or administrative zoning. Because the classes are defined by physical structure alone, an industrial estate and a dense low-rise residential district with similar building density and height can fall in the same LCZ, and are expected to exhibit similar thermal behaviour. This decoupling from land-use categories is precisely

[15] (Zhou & Mutani, 2026)

[16] (Mutani & Todeschi, The effects of green roofs on outdoor thermal comfort, urban heat island mitigation and energy savings, 2020)

[17] (Clinton & Gong, 2013)

[18] (Stewart & Oke, 2012)

what makes the scheme transferable between cities and suitable as an explanatory variable in cross-city SUHI models.[19]

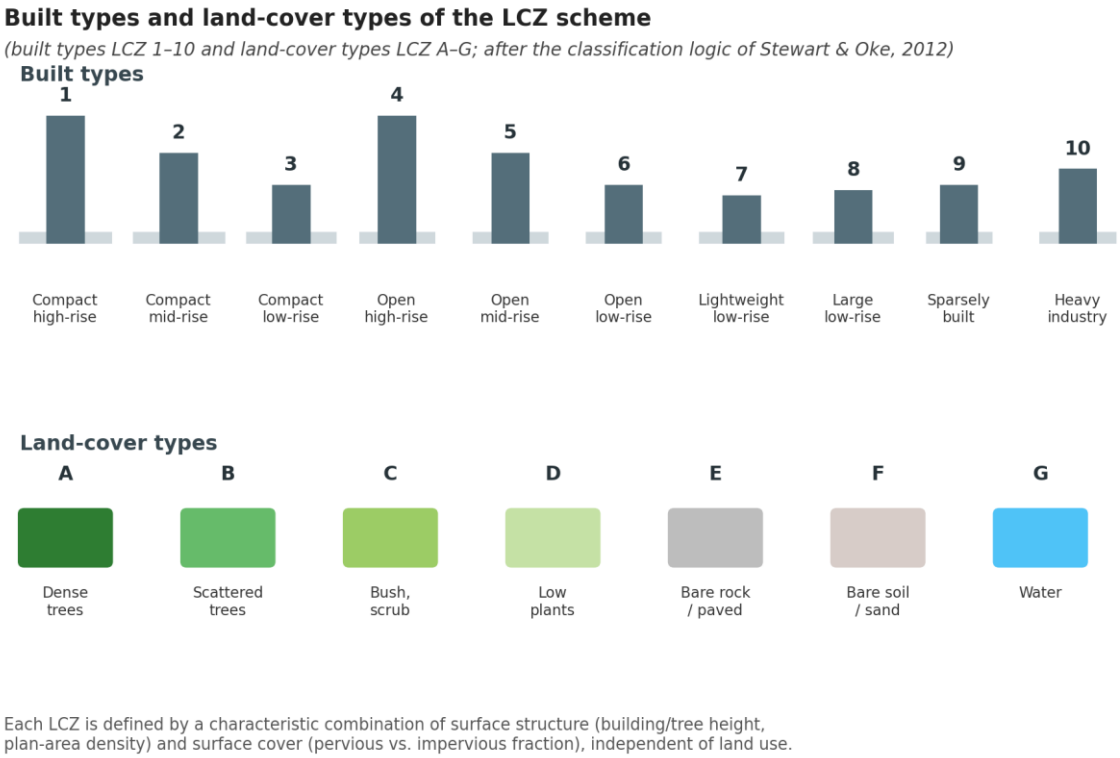


Figure 1.2. Schematic overview of the seventeen LCZ classes (ten built types, seven land-cover types) by Author.

The LCZ scheme underpins the World Urban Database and Access Portal Tools (WUDAPT) initiative[20], under which a globally consistent, Landsat-derived LCZ map was produced using a Random Forest classifier trained on crowd-sourced Level-0 training areas.[21] A refined version of this classification, incorporating a morphology-aware Gaussian post-processing filter to reduce salt-and-pepper noise while preserving linear features such as rivers and arterial roads, was

[19] (Bechtel, et al., 2019)

[20] WUDAPT stands for the World Urban Database and Access Portal Tools, a community-driven initiative that supplies the standardised training data underlying the global LCZ products used in this thesis.

[21] (Demuzere, Bechtel, Middel, & Mills, 2020)

released as Demuzere, Kittner and Bechtel[22] and is the dataset used throughout this thesis (see Section 4.1 for the adaptation of this product to the 100 m Turin study grid).

Because it is derived from a global model rather than a city-specific classification, this product[23] necessarily trades local accuracy for consistency and reproducibility; the authors report an overall global classification accuracy of 71.2% and note that accuracy for individual cities can deviate from this figure, particularly for classes that occupy a small mapped area or exhibit mixed morphology. Section 4.3 discusses how this global classification relates to the locally computed morphological indicators derived for Turin, a computationally lighter alternative to the bespoke rule-based classification sometimes attempted in single-city studies.

1.4 Machine learning approaches to attributing morphological and climatic drivers of SUHI

Early SUHI–morphology studies relied primarily on linear correlation and ordinary least-squares regression to relate candidate predictors to observed SUHI intensity. These methods remain informative and are retained in this thesis as a first screening step (Section 5.2), but they assume additive, linear relationships and are known to understate the importance of predictors that interact non-linearly with others. Ensemble tree-based methods, and Random Forest in particular,[24] have become the dominant tool for SUHI driver attribution over the past decade because they accommodate non-linear effects and variable

[22] (Demuzere, Kittner, & Bechtel, A global map of local climate zones to support earth system modelling and urban-scale environmental science, 2022)

[23] (Demuzere, Kittner, & Bechtel, A global map of local climate zones to support earth system modelling and urban-scale environmental science, 2022)

[24] (Breiman, 2001)

interactions without requiring the predictor set to be pre-specified in functional form, and because they naturally handle the mixed continuous/categorical variable sets typical of morphology-plus-climate SUHI models.[25]

Indicator	**Definition**	**Typical role in SUHI models**	**Key reference**
Sky View Factor (SVF)	Visible sky hemisphere fraction from a point located at street level	Governs long-wave radiation loss at night time.	Oke (1982)
Building Coverage Ratio (BCR)	Area fraction covered by building footprints	Proxy for impervious surface / thermal mass fraction	Stewart & Oke (2012)
Height-to-Width ratio (H/W)	Mean building height divided by street canyon width	Shapes canyon geometry, radiation trapping	Oke (1982)
Impervious Surface Fraction (ISF)	Fraction of grid cell covered by impervious materials	Reduces cooling through evapotranspiration	Bechtel et al. (2019)
NDVI	Normalised Difference Vegetation Index	Vegetation density, evapotranspiration proxy	Peng et al. (2012)
NDBI	Normalised Difference Built-up Index	Built-up surface extent	Zhang et al. (2022)
Albedo	Surface reflectance	Controls absorbed short-wave radiation	Liang (2001)

Table 1.1. Morphological and spectral indicators commonly used as SUHI predictors in the reviewed literature; the full set computed for Turin is described in Chapters 3–4.

A well-known limitation of Random Forest and other ensemble methods is their reduced interpretability relative to a regression coefficient: a trained forest can predict SUHI accurately without revealing, in a form usable by planners, *why* a given cell is predicted to be hot. SHapley Additive exPlanations (SHAP), based

[25] (Zhang, Li, & Liu, 2022); (Zhou & Mutani, 2026)

on cooperative game theory, address this by attributing each prediction to the individual contribution of each input feature in a way that sums exactly to the model output for that observation, and that is consistent across features of different scale and type.[26] This property is what allows the seasonal and annual SHAP analyses in Chapter 5 to rank morphological and climatic drivers of SUHI on a common footing, offering a more rigorous alternative to the "mean decrease in impurity" importance scores native to the Random Forest algorithm itself.

Because morphological indicators are frequently correlated with one another by construction, a cell with a high building coverage ratio will typically also have a low sky view factor and a high impervious surface fraction. Additionally, multicollinearity screening via the Variance Inflation Factor is treated in this thesis as a mandatory pre-modelling step (Section 5.1), following standard practice in applied SUHI regression studies.

1.5 Consequences of Surface Urban Heat Islands

Beyond its value as a measurable physical phenomenon, SUHI carries direct consequences for energy systems, public health and the urban water cycle. That is why it has become a recurring concern in urban planning research rather than a purely academic curiosity. As impervious surfaces store and re-emit more heat than the vegetated or pervious surfaces they replace, buildings within affected areas typically require more energy for space cooling during warm periods. In turn increasing both household energy expenditure and peak

[26] (Lundberg & Lee, 2017)

demand on local power infrastructure, with the additional heat rejected by cooling systems reinforcing the very warming it responds to.[27]

The public-health dimension is at least as significant. Higher surface and near-surface temperatures don't straightforwardly raise air temperature, though the connection to increased outdoor thermal stress remains consistent. Prolonged exposure to hot urban environments has been linked to increased heat-related morbidity, with older residents, children and socio-economically disadvantaged groups disproportionately affected. A recent micro-scale assessment of the heat–health nexus in Turin specifically found that vulnerability to heat is shaped jointly by socio-economic factors and by built-environment characteristics such as building density and surface materials, underlining that SUHI mapping has a direct bearing on equity as well as comfort.[28]

The third consequence is the effect on urban water cycle caused by replacement of pervious surfaces with impervious surface. It reduces the water infiltration and accelerates the surface water runoff while limiting the ground water recharge. It also reduces evapotranspiration that transfers part of the incoming solar energy away as latent heat. This results in retaining the energy at the surface contributing to the warming effect. Decades of tracking impervious-surface growth point to something beyond a simple additive effect: land surface temperature shifts more than reduced vegetation and water loss alone would account for, a pattern that fits with the radiative and thermal-admittance dynamics laid out in Section 1.[29]

[27] (Vujovic, Haddad, Karaky, Sebaibi, & Boutouil, 2021)

[28] (Ellena, et al., 2023)

[29] (Xu, Lin, & Tang, 2013)

Taken together, these three consequences set up the two remaining strands of this literature review: identifying which morphological and climatic variables drive SUHI intensity (Sections 1.3–1.4), and translating that understanding into concrete mitigation and compensation measures (Section 1.6).

1.6 Nature-based solutions and environmental compensation frameworks for UHI mitigation

Nature-based solutions (NBS), an umbrella term covering interventions such as tree planting, green roofs, permeable paving and urban wetlands - are among the most extensively evidenced approaches to reducing local urban temperatures. A systematic review of the empirical literature found that parks were, on average, cooler than their surrounding built environment by an amount equivalent to just under one degree Celsius during the daytime on average,[30] with larger, tree-dense parks generally showing a stronger cooling effect.

Trace Italy's approach to environmental mitigation and compensation programmes and projects of the past and it leads to EU law. The Environmental Impact Assessment (EIA) and Strategic Environmental Assessment (SEA) Directives were changed nationally through Legislative Decree No. 152/2006. On this basis, Piedmont added its own layer via Regional Law No. 13/2023. Turin's Metropolitan City built further on this foundation in 2023 which created the Catalogue of Regeneration and Environmental Compensation Interventions (C.I.R.C.A) through its General Territorial Metropolitan Plan (PTGM). The catalogue works as a structured inventory of compensation measures, meant to encourage public bodies and project proponents toward interventions that

[30] (Bowler, Buyung-Ali, Knight, & Pullin, 2010)

actually fit the site's environmental context.[31] The catalogue is organised around five objectives:

- Improving the quality of environmental components (air, water, soil);
- Expanding green and blue infrastructure and ecological connectivity;
- Regenerating degraded areas;
- Increasing resilience and adaptability to climate change; and
- Preserving and increasing biodiversity and the naturalness of the territory.

The catalogue includes eight categories as: ecological renaturalisation, forestation, river and hydromorphological requalification, biodiversity protection, pollution reduction, green infrastructure, sustainable urban drainage, and measures specifically designed for climate change and urban heat islands.[32] This follows the idea that heat mitigation works best when different types of interventions act together rather a single intervention acting individually. For example, a green roof reduces the surface temperature by shading and evapotranspiration, slows the stormwater runoff, and provides a dwelling for urban wildlife to live. The thermal, hydrological and ecological outcomes discussed in Section 1.5 represents exactly this idea.

Each catalogued intervention is documented in a standardised technical sheet describing its function, constituent works, illustrative precedents and supporting references,[33] a format that this thesis draws on directly when linking LCZ-specific mitigation recommendations to catalogue entries in Section 6.3. The explicit link between a spatial classification of urban structure and a spatial inventory of feasible interventions remains largely theoretical in the current literature;

[31] (Barbieri, Bottero, Caprioli, Cassatella, & Datola, 2026)
[32] (Barbieri, Bottero, Caprioli, Cassatella, & Datola, 2026)
[33] (Barbieri, Bottero, Caprioli, Cassatella, & Datola, 2026)

connecting the two systematically, using the LCZ map as the spatial key, is one of the contributions this thesis attempts to make.

1.7 Research gap and objectives of this thesis

Taken together, the literature reviewed in Sections 1.1–1.6 establishes four things clearly: that SUHI can be measured reliably and repeatably from Landsat-class thermal imagery; that the Local Climate Zone scheme provides a transferable, structure-based framework for stratifying that measurement; that machine learning, and Random Forest with SHAP in particular, is now the standard approach for attributing SUHI to its morphological and climatic drivers without over-simplifying their interactions; and that a well-developed, evidence-based catalogue of mitigation and compensation interventions already exists for the Turin context specifically. Few studies in the existing literature integrate all four of these elements within a single reproducible, city-wide workflow. This thesis combines a globally consistent LCZ classification with locally derived morphological indicators. Parallel z-score and absolute SUHI indices are computed across four representative seasonal dates, analysing the full annual thermal cycle. A random forest-SHAP attribution is then used to identify the dominant drivers at each seasonal scale and across the combined annual dataset. And also links those drivers directly through the LCZ classification to the C.I.R.C.A. catalogue of compensation measures actually available to Turin's planning authorities. Accordingly, the objectives of this thesis are to:

1. Derive Land Surface Temperature, SUHII (z-score) and SUHI (rural-reference) indices for Turin across four representative seasonal dates in a climatologically typical year;

2. Characterise Turin's urban morphology at 100 m resolution using the Demuzere et al. (2022) LCZ classification alongside locally computed morphological indicators;

3. Quantify the relationship between variables and SUHII/SUHI, through correlation analysis and a Random Forest model, interpreted via SHAP models, separately by season and for the full year;

4. Identify the variables most consistently associated with high SUHI/SUHII across seasons; and

5. Translate these findings into season specific, SHAP derived intervention thresholds and a spatially differentiated feasibility framework. Linking the resulting residual thermal burden to compensation measures available through the C.I.R.C.A. catalogue of Metropolitan City of Turin.

Chapter 2 introduces Turin as the case study and describes the data sources and workflow used to meet these objectives; Chapters 3–5 implement the indicator derivation, LCZ characterisation, and modelling steps in turn; Chapter 6 develops the mitigation and compensation analysis; and Chapter 7 discusses the results and their implications for urban planning in Turin.

Chapter 2- Turin as a Case Study: Urban Context, Monitoring Network and Research Workflow

2.1 Geographical and climatic context of Turin

Turin which is the administrative capital of Piedmont in north-western Italy, sits at the western edge of the Po Valley, right where the land begins to climb into the Alps. Local studies of the city's climate keep coming back to this setting: Turin sits inside a broad basin which is largely surrounded by high ground on most sides, and that enclosure limits how freely air can circulate through the city. The result is a tendency toward stable, settled weather patterns, most noticeably through the summer months.[34] This combination of urban density and constrained atmospheric circulation is one of the reasons Turin has attracted sustained research attention as a case study for Surface Urban Heat Island analysis.

The consequences of SUHI introduced in Chapter 1 are not abstract for Turin specifically. The same micro-scale assessment of the heat–health nexus discussed there found that vulnerability to heat within the city is unevenly distributed, shaped jointly by socio-economic conditions and by built-environment characteristics such as building density and surface materials,[35] the same morphological factors this thesis investigates as SUHI drivers. This finding motivates a spatially explicit approach to SUHI analysis, since city-wide averages would obscure exactly the neighbourhood-level differences that matter most for targeting mitigation where it is most needed.

[34] (Mutani, Scalise, Sufa, & Grasso, 2024)

[35] (Ellena, et al., 2023)

The city's climate is classified as humid subtropical under the Köppen–Geiger scheme (Cfa), characterised by cold, often foggy winters and warm summers with a marked seasonal temperature range. July is consistently the warmest month and January the coldest, with a secondary rainfall peak typically occurring in spring and autumn. For the 2025 typical year selected for this thesis, mean monthly air temperature across the ARPA network ranges from 3.91 °C in January to 24.69 °C in June, giving an annual mean of 14.25 °C; the full station-by-station breakdown and the selection procedure behind this choice of year are set out in Section 3.1.

This pronounced seasonality has a direct bearing on the design of the remote-sensing component of this thesis. Because vegetation phenology, solar elevation and background air temperature all vary substantially across the year, a single summer acquisition would characterise only one point in a much more variable annual cycle. This is the rationale, introduced in Chapter 1, for deriving Land Surface Temperature and the associated indicators from four representative seasonal dates within a single climatologically typical year, in preference to a single scene or a multi-year summer-only series (Section 3.1).

2.2 Ground-based monitoring network: ARPA regional stations and the Via della Consolata urban station

Two distinct tiers of ground-based meteorological monitoring underpin this thesis. The first is the regional network operated by ARPA Piemonte, comprising several hundred automatic stations distributed across the region at an average density of roughly one station per 100 km², recording temperature, precipitation, humidity, wind and, at a subset of sites, solar radiation. Records from this network

for the years 2021–2025 provide the basis for identifying a climatologically typical year, described in Section 3.1.

The second tier is the Via della Consolata station, an urban meteorological site located within Turin's historic centre. Unlike the wider ARPA network, which mixes urban, peri-urban and rural sites, Via della Consolata specifically characterises atmospheric conditions within the built-up city, and its air temperature, relative humidity, wind speed and direction, and solar radiation records are used in this thesis to describe the meteorological context of each Landsat acquisition (Section 3.2.3).

Alongside this, third complementary role is played by the Moncalieri Baducchi weather station, located south of the city and used as the rural reference weather station for the calculation of absolute SUHI (Section 3.3). The weather station is chosen after overlaying its coordinates on the Local Climate Zone map which confirms that Baducchi falls predominantly within LCZ D (low plants), consistent with the low-density, vegetated reference conditions that the LCZ scheme itself associates with rural benchmarking.[36]

2.3 Overview of data sources: satellite imagery, the Demuzere LCZ dataset and Geoportale Torino

Beyond the ground-based networks described above, three further data sources structure the analytical workflow of this thesis. The first is Landsat 8/9 Collection 2 Level-2 imagery, processed in Google Earth Engine to derive the Land Surface Temperature, SUHI/SUHII and ancillary spectral indicators introduced in Chapter 1 and detailed in Section 3.2. Landsat's 30 m nominal spatial resolution and multi-decade continuity make it well suited to city-scale,

[36] (Stewart & Oke, 2012)

repeatable SUHI assessment, at the cost of the coarser thermal detail associated with higher-resolution but less consistently archived sensors.

The second is the Demuzere et al. (2022) global Local Climate Zone classification, used throughout this thesis to characterise Turin's urban structure at 100 m resolution.[37]

The cartographic data is obtained from the Geoportale della Città di Torino which is the municipal geoportal maintained by the Comune di Torino. It provides access to the Carta Tecnica Comunale and other associated vector datasets covering the city. This dataset provides the building footprints, heights and street geometries from which the morphological indicators used in this thesis (building coverage ratio, height-to-width ratio, sky view factor, pervious and impervious surface fraction, and green coverage ratio) are computed in QGIS, as detailed in Section 4.2.

2.4 Methodological workflow: from typical-year selection to mitigation strategies

The five data sources introduced in this chapter feed into a single, sequential analytical workflow, summarised in Figure 2.1. The workflow begins with the identification of a climatologically typical year from the ARPA Piemonte record and the selection of four representative Landsat scenes for that year, one per season. Each scene is then processed in Google Earth Engine to derive Land Surface Temperature, SUHII and SUHI, and a set of ancillary spectral indicators.

[37] (Demuzere, Kittner, & Bechtel, A global map of local climate zones to support earth system modelling and urban-scale environmental science, 2022)



Figure 2.1. Five-stage methodological workflow followed in this thesis, from typical-year and scene selection through to LCZ-specific mitigation and compensation.

In parallel, Turin's urban structure is characterised by combining the Demuzere Local Climate Zone classification with morphological indicators computed from Geoportale Torino data, providing the explanatory variables used in the fourth stage of the workflow: a Random Forest model, interpreted through SHAP and preceded by Pearson correlation and multicollinearity screening, run separately for each season and for the full year. Dependence curves derived from this model extract season-specific turning points for each actionable predictor. These turning points carry into the fifth and final stage as season-specific intervention thresholds for NDVI, ISF, and Albedo. An LCZ-based feasibility screening flags cells in the built grid that are structurally not feasible for any direct

intervention. The residual thermal burden on these cells is directed to programmes through C.I.R.C.A. compensation catalogue of the Metropolitan City of Turin.

This five-stage structure maps directly onto the remaining chapters of the thesis: Chapter 3 covers the first two stages (typical-year selection and indicator derivation), Chapter 4 covers the third (LCZ and morphology), Chapter 5 covers the fourth (statistical modelling), and Chapter 6 covers the fifth (mitigation and compensation), before Chapter 7 draws the results together.

Chapter 3- Selecting a Representative Year and Deriving Remote-Sensing and Climate Indicators

3.1 Selection of a climatologically typical year

Before any of the remote-sensing indicators introduced in Chapter 2 can be computed, a single representative year has to be identified so that the four seasonal Landsat scenes selected in Section 3.2 are not drawn from an atypically hot or cold twelve months. Following the deviation-minimisation procedure used in earlier urban-scale climate and energy modelling for Turin,[38] a climatologically typical year is derived directly from the ARPA Piemonte monitoring network, without relying on an external reference dataset. The dataset incorporates mean monthly air temperatures spanning from 2021 to 2025 across fifteen ARPA province stations (Table 3.1). These stations are positioned at elevations from 192 to 619 meters above sea level and extending up to 17.5 kilometres from the urban centre.

[38] (Mutani, Cristino, & Bullita, The Urban Heat Island of the Metropolitan City of Turin: Strategies for a Sustainable Urban Planning, 2018)

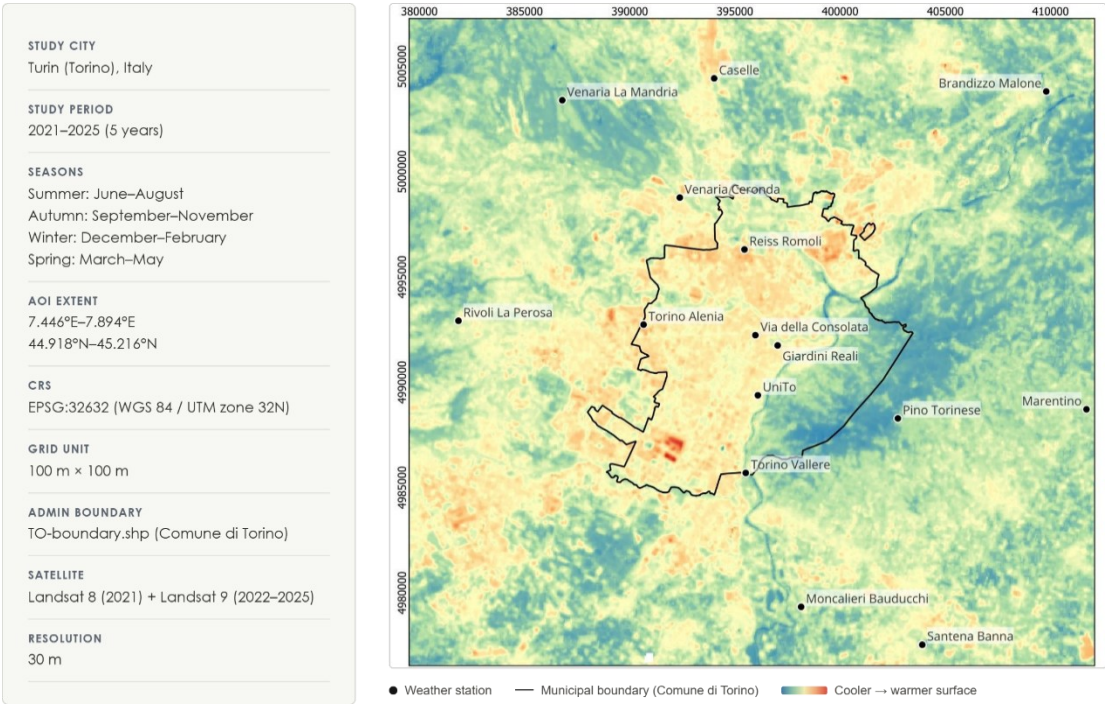


Figure 3.1. Study area, Landsat coverage and ARPA weather-station network used for typical-year selection and for the remote-sensing indicators derived in this chapter.

For each station and year, the twelve-monthly means are collapsed into a single annual value; the fifteen station values are then aggregated spatially to yield one annual mean temperature per year, and the five resulting annual means are pooled to produce the five-year period mean of 14.10 °C. The candidate typical year is identified as the one whose annual mean sits closest to that period mean in absolute terms. Table 3.2 presents the full picture: 2021 occupies the outermost position relative to the period mean, 2022 registers as the highest-temperature year across the record, while 2024 and 2025 cluster near the centre, with deviations of 0.09 °C and 0.15 °C respectively — a separation of just 0.06 °C between the two leading candidates.

This narrow margin means that both years are statistically valid typical-year candidates, and the choice between them is resolved primarily by satellite data availability, since the temperature criterion alone cannot separate the two years with confidence. Land Surface Temperature retrieval requires each Landsat

scene to have less than 10 % cloud cover over the Turin area of interest (7.446°E–7.894°E, 44.918°N–45.216°N); a review of the USGS Earth Explorer archive shows that 2025 offers substantially more cloud-free Landsat 8/9 acquisitions than 2024 for both the summer and winter windows. Using a year with few suitable scenes would force multi-year compositing and reintroduce the inter-annual variability that the typical-year selection is intended to remove, so 2025 is adopted as the typical year for the remainder of this thesis, consistent with the imagery-quality considerations that also motivated the move away from 2024 at the outset of this study.

Station	**ARPA ID**	**Municip.**	**Lat. N**	**Long. E**	**Alt. (m)**	**Dist. (km)**	**Active since**
Giardini Reali	446	Torino	45.0717	7.6919	239	0.123	2004
Reiss Romoli	S3869	Torino	45.1125	7.6709	270	4.454	2003
Torino Alenia	S4294	Torino	45.0796	7.6107	320	5.173	2005
Torino Vallere	249	Torino	45.0170	7.6739	222	4.096	2001
Via della Consolata	S3447	Torino	45.0759	7.6784	290	0.484	2003
UniTo	—	Torino	45.0503	7.6804	254	1.737	2004
Brandizzo Malone	373	Brandizzo	45.1820	7.8519	192	17.525	2005
Caselle	S2896	Caselle T.	45.1854	7.6509	300	12.15	2003
Marentino	—	—	45.0464	7.8789	345	16.07	1995
Moncalieri Bauducchi	273	Moncalieri	44.9601	7.7086	226	11.206	1993
Pino Torinese	120	Pino Torinese	45.0413	7.7652	619	6.72	1988
Rivoli La Perosa	S3948	Rivoli	45.0799	7.4989	362	12.843	2004
Santena Banna	—	—	44.9447	7.7819	238	16.31	1998
Venaria La Mandria	S4587	Druento	45.1750	7.5592	337	13.465	2006
Venaria Ceronda	256	Venaria	45.1342	7.6313	253	7.881	1997

Table 3.1. ARPA Piemonte weather stations used to identify the 2025 typical year.

Year	Annual mean T (°C)	|Deviation| (°C)	Status
2021	13.27	0.83	—
2022	14.65	0.55	Hottest year in record
2023	14.33	0.23	—
2024	14.01	0.09	Closest statistically; not selected (scene availability)
2025	14.25	0.15	Selected as typical year
Period mean (2021–2025)	14.10	—	—

Table 3.2. Annual temperature deviation analysis, 2021–2025 (ARPA Piemonte, fifteen-station spatial average). Deviation is given as the absolute difference between the annual mean and the 14.10 °C period mean.

The twelve-monthly values that were averaged into each annual mean above are set out in Table 3.3 for the selected 2025 typical year, so that its seasonal shape can be checked directly against the broader five-year pattern reported in Table 3.2, alongside the single annual figure already discussed.

Month	Mean air temperature (°C), fifteen-station spatial average
January	3.91
February	5.92
March	9.79
April	14.03
May	17.40
June	24.69
July	23.74
August	23.97
September	18.71
October	13.63
November	7.36
December	5.80
Annual mean	14.25

Table 3.3. Monthly mean air temperature for the selected 2025 typical year, fifteen-station spatial average (ARPA Piemonte).

3.2 Satellite scene selection and pre-processing

3.2.1 Overview of the Google Earth Engine processing workflow

The scene-selection procedure described above is one stage of a single custom Google Earth Engine script, developed with reference to the Turin-focused GEE workflow of Zhou and Mutani[39] and applied identically to each of the four seasonal scenes and documented in full in Appendix A. Table 3.4 summarises the seven stages of this script, from the initial selection of the Landsat image through to the export of the raster layers used in the remainder of this thesis; each stage is described in more technical detail in the sections that follow.

Stage	Description
1. Image selection	Filter the merged Landsat 8/9 Collection 2 Level-2 catalogue by the Turin AOI and seasonal date window, print the ranked list of every candidate scene under 10% cloud cover, and select the least-cloudy one by default (or a manually specified acquisition date, when set).
2. Scaling factors	Apply the Collection 2 scale factors to convert digital numbers to surface reflectance (optical bands) and at-sensor brightness temperature (thermal band).
3. Spectral indices	Calculate NDVI, NDBI, NDWI and NDMI from the scaled optical bands.
4. Emissivity chain	Derive PVI from NDVI, then compute Land Surface Emissivity from PVI.
5. LST retrieval	Correct the Band 10 brightness temperature for emissivity using the Avdan–Jovanovska algorithm to obtain Land Surface Temperature in °C.
6. SUHII mapping	Standardise the LST raster into the SUHII z-score using the AOI-wide mean and standard deviation.
7. Export	Export all raster layers as GeoTIFF files (30 m resolution in EPSG:32632) for the QGIS-based analysis of Chapter 4.

Table 3.4. Overview of the Google Earth Engine script used to extract and process each seasonal Landsat scene (full script in Appendix A).

[39] (Zhou & Mutani, 2026)

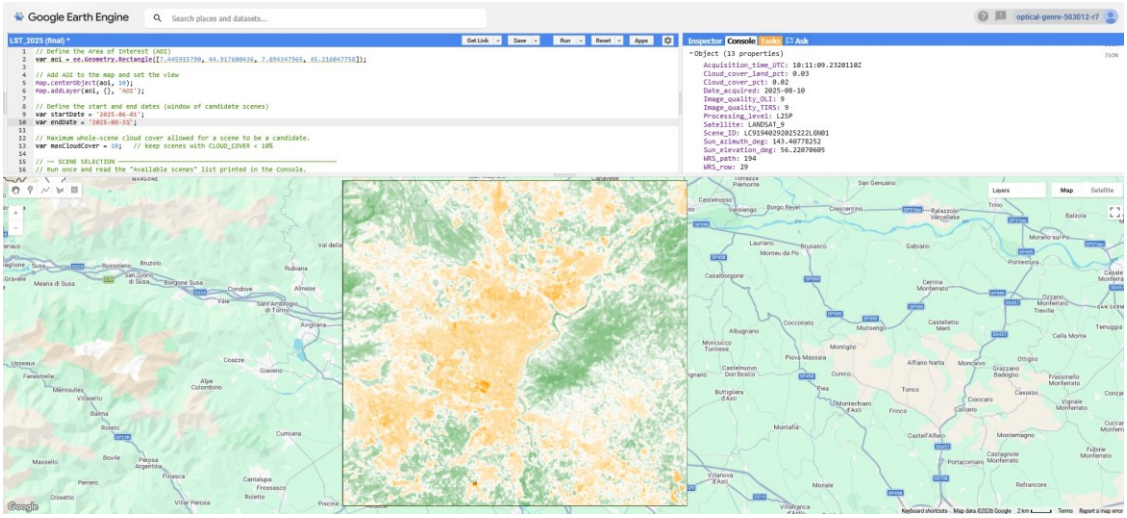


Figure 3.2. The Google Earth Engine interface with Summer 2025 imagery query in a single view, with the Area of Interest boundary and the seasonal date parameters applied.

Figure 3.2 illustrates the first of these seven stages for the summer 2025 acquisition. The Area of Interest is defined in the script as a rectangle spanning the coordinates given in Section 3.1, and the map view is centred on this polygon before the seasonal date window (1 June to 31 August 2025) and the maximum cloud-cover threshold are applied as filters to the merged Landsat 8 and 9 Collection 2 Level-2 image collection, sorted by ascending cloud cover. The script does not commit silently to the top-ranked scene; it first prints the full candidate list to the Console — acquisition date, whole-scene and land-only cloud fraction, satellite platform and scene identifier for every image under the 10 % threshold — so that the ranking can be checked before a scene is accepted. For the summer, winter and spring windows, the automatic least-cloudy selection was accepted directly from this ranked list, with the pickDate variable left unset. For the autumn window, this initial pass returned eight candidate scenes under the cloud-cover threshold; the 5 October 2025 scene was chosen from this list and its acquisition date was then hardcoded into the pickDate variable, fixing the extraction to that specific scene and avoiding any

risk of a different scene being selected automatically on a later run, at a reported cloud fraction of 0.2 % consistent with Table 3.5. Following automated filtering and sorting, the Console panel on the right of Figure 3.2 displays the metadata for the selected scene, detailing parameters including Scene ID, acquisition time, sun geometry, WRS path and row, and cloud-cover percentages across both the entire scene and its land area. For this capture, the data verifies a minimal cloud cover of 0.02 percent, in agreement with Table 3.5 and attributes the image (ID LC91940292025222LGN01) to the Landsat 9 satellite. The same candidate-list inspection is repeated for each of the remaining three seasonal windows, with the winter and spring scenes accepted automatically in the same way as the summer scene shown here, before the script proceeds to the scaling-factor and spectral-index stages summarised in Table 3.4.

The Turin-specific application of this workflow — including the exact scenes selected for each season and the resulting cloud-cover values — is described in the remainder of this section, while the emissivity and Land Surface Temperature chain is set out in Section 3.3 and the SUHII and SUHI calculations in Section 3.5.

Landsat 8 and 9 Collection 2 Level-2 surface reflectance and surface temperature products are retrieved through Google Earth Engine at their native 30 m resolution and reprojected to EPSG:32632 (WGS 84 / UTM Zone 32N) over the Turin area of interest. Four seasonal windows are defined following the meteorological seasons adopted throughout this thesis — summer (June–August), autumn (September–November), winter (December–February) and spring (March–May) — and, within each window, candidate scenes are screened for cloud cover below 10 %. Selected scenes are further cross-checked against ARPA hourly records at the Torino Caselle station, so that only images acquired under stable, largely cloud-free and low-wind conditions are retained,

consistent with the general requirement that surface-temperature retrievals be drawn from radiatively undisturbed atmospheric conditions.[40]

The four scenes selected for the 2025 typical year are summarised in Table 3.5. The summer scene, acquired on 10 August 2025, has a negligible cloud fraction of 0.02 %; the winter scene of 6 February 2025 is the cloudiest of the four at 4.6 %; the spring scene of 3 March 2025 has 1.99 % cloud cover; and the autumn scene of 5 October 2025 has the lowest cloud fraction of the four dates, at 0.2 %. All four dates are treated as the working set of seasonal scenes for the remainder of the thesis.

Season	Scene date	Cloud cover
Summer	10 August 2025	0.02%
Winter	6 February 2025	4.6%
Spring	3 March 2025	1.99%
Autumn	5 October 2025	0.2%

Table 3.5. Landsat 8/9 scenes selected for the 2025 typical year, by season.

Radiometric preprocessing follows the standard Collection 2 scaling coefficients: optical bands are converted to surface reflectance as DN × 0.0000275 − 0.2, and the thermal band is converted to at-sensor brightness temperature in kelvin as DN × 0.00341802 + 149.0. Per-pixel cloud and cloud-shadow screening is nominally available through the QA_PIXEL bits 3 and 4 (the CFMask flags); in the Earth Engine processing carried out for this thesis, however, the cloud-masking function is present in the script but is not invoked, so quality control is enforced entirely at the scene-selection stage, through the 10 % scene-level cloud-cover

[40] (Oke, 1982)

threshold described above, without per-pixel masking.[41] All derived layers are exported as GeoTIFF at 30 m resolution in EPSG:32632 for the subsequent morphological analysis carried out in QGIS (Section 4.2).

3.3 Land Surface Temperature retrieval

Land Surface Temperature is retrieved with the four-step, emissivity-corrected single-channel algorithm of Avdan and Jovanovska,[42] implemented in Google Earth Engine following the open-source workflow of Ermida et al.[43] and adapted from the reference script used in related downscaling work on Turin.[44] The procedure calculates the Normalized Difference Vegetation Index (NDVI) via the red and near-infrared bands, using this metric to determine the Proportion of Vegetation (PVI) by squaring NDVI after rescaling it within its local boundaries. Derived from PVI through a linear adjustment bounded between 0.986 and 0.990, Land Surface Emissivity is subsequently applied to the Band 10 at-sensor brightness temperature to produce Land Surface Temperature values in degrees Celsius (Table 3.6).

Step	Formula	Range / unit
1. NDVI	(NIR − RED) / (NIR + RED)	−1 to +1
2. PVI	((NDVI − NDVImin) / (NDVImax − NDVImin))²	0 to 1
3. Land Surface Emissivity (ε)	0.004 × PVI + 0.986	0.986–0.990
4. Land Surface Temperature	T_b / [1 + (0.00115 × T_b / 1.438) × ln(ε)] − 273.15	°C

Table 3.6. Land Surface Temperature retrieval process referred from Avdan and Jovanovska (2016).

[41] This is recorded here as a methodological limitation: redoing the LST extraction with per-pixel masking enabled would change the underlying rasters used throughout this thesis, so the limitation is noted here. It is revisited among the sources of uncertainty discussed in Chapter 7.

[42] (Avdan & Jovanovska, 2016)

[43] (Ermida, Soares, Mantas, Göttsche, & Trigo, 2020)

[44] (Zhou & Mutani, 2026)

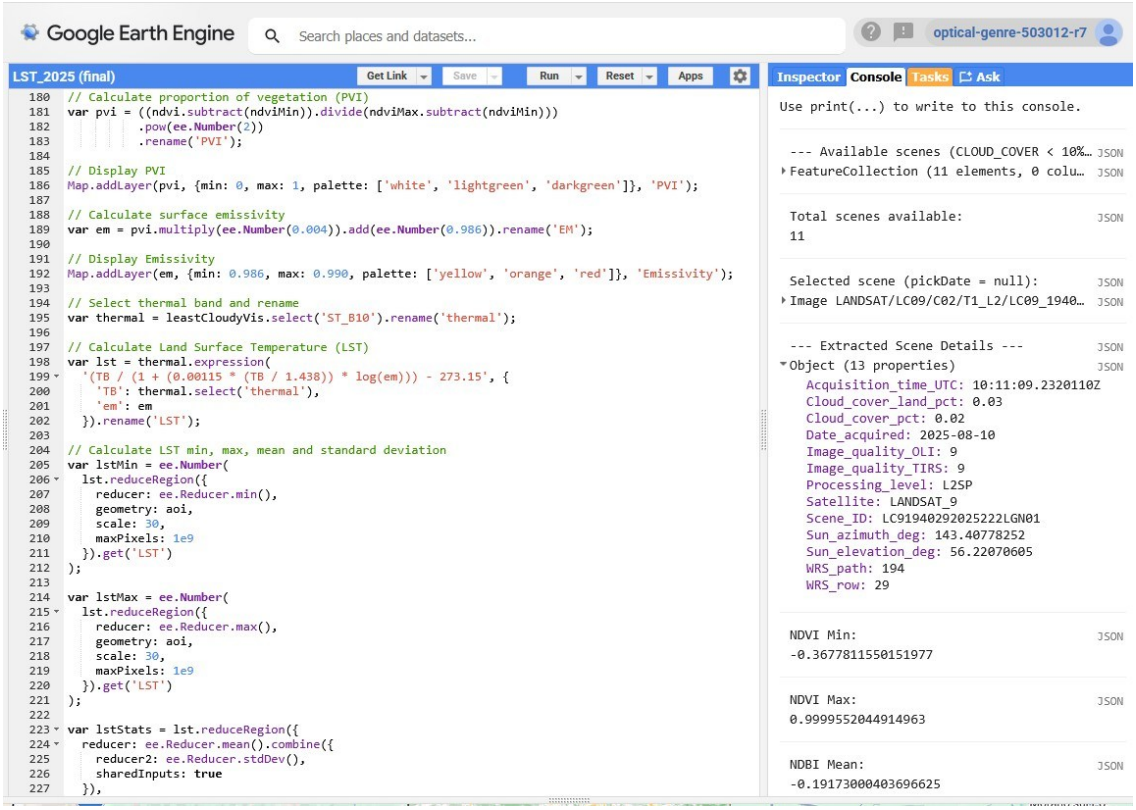


Figure 3.3. Google Earth Engine script and Console for the PVI, emissivity, Land Surface Temperature stages, for the summer 2025 scene.

Figure 3.3 shows the corresponding stage of the script for the summer 2025 scene. The code computes PVI from the NDVI extremes taken over the AOI, converts PVI into surface emissivity, and applies that emissivity to correct the Band 10 brightness temperature into LST. The script also derives the AOI-wide LST minimum, maximum, mean and standard deviation via separate reduce Region calls, the last two of which feed the SUHII standardisation described in Section 3.5. The Console lists the total number of candidate scenes (11), the metadata of the selected scene (Landsat 9, acquired 10 August 2025, 0.02% cloud cover), and the beginning of the printed index diagnostics -an NDVI range of 0.368 to 1.000 and an NDBI mean of −0.192 for this scene.

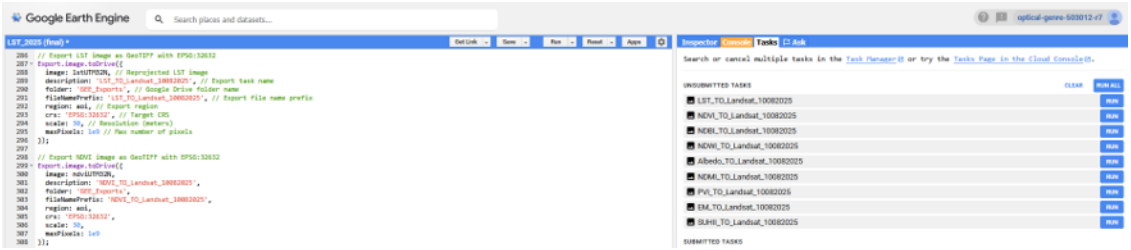


Figure 3.4. Google Earth Engine Tasks panel showing the unsubmitted GeoTIFF export tasks for the nine raster layers listed in Table 3.7, with the LST export block's parameters visible in the script on the left.

Figure 3.4 captures the Tasks panel with the accompanying code block specifying, for each layer, the full set of export parameters: the target Google Drive folder and file name, the export region (the AOI defined in Section 3.1), the coordinate reference system (EPSG:32632), the output resolution (30 m), and the maximum permitted pixel count. A layer and date specific identifier travels with each task, for example, LST_TO_Landsat_10082025 — when run this enters the queue to the resulting GeoTIFF filename. By submitting all nine layers together under a shared parameter set keeps the given scene's output on an identical spatial grid, that eases both the QGIS import and the zonal-statistics workflow described in Chapter 4. Each queued task requires a manual click from this panel to begin running.

Figure 3.5 shows the resulting Land Surface Temperature raster for all four seasonal 2025 scenes. The summer layer displays its highest values over the urban core and the built-up fringe to the north-east, reaching approximately 63.7 °C at the mid-morning Landsat overpass, and its lowest over the Po and Dora Riparia river corridors and the surrounding agricultural belt to the south-west, down to approximately 29.7 °C, consistent with the exact AOI-wide extremes reported directly from Google Earth Engine in Figure 3.3. The spring and autumn layers display values between the summer and winter extremes, spanning approximately −4.2 °C to 31.0 °C in March and approximately 4.2 °C to 39.8 °C

in October, with the autumn scene running noticeably warmer overall than the spring one despite the two dates sitting at comparable points in the annual temperature cycle. The winter layer displays the most compressed range of the four, spanning approximately −7.4 °C in the coldest rural and hillside areas to approximately 22.3 °C in the most built-up sections of the city, and the built-up/vegetated pattern that dominates the summer raster becomes far less pronounced, consistent with the surface-energy-balance mechanisms discussed in Chapter 1.

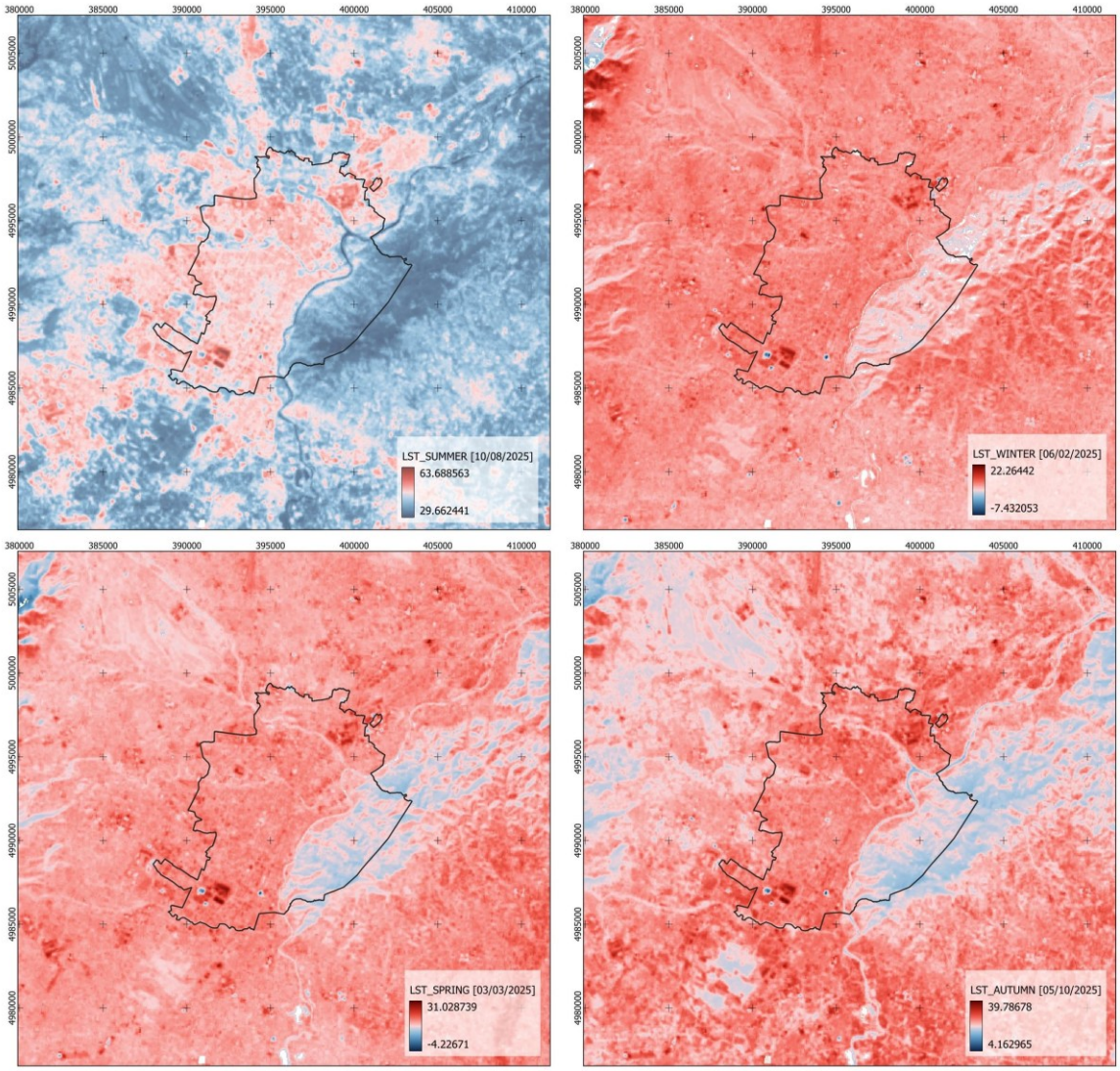


Figure 3.5. Land Surface Temperature retrieved for all four seasons in 2025.

3.4 Ancillary spectral and biophysical indicators

Beyond the Land Surface Temperature, seven additional layers are derived from the same Landsat 8/9 imagery, serving both as predictors in the Random Forest models of Chapter 5 and as intermediate quantities within the emissivity chain described above. The Normalised Difference Built-up Index, NDBI = (SWIR1 − NIR) / (SWIR1 + NIR), and Normalised Difference Water Index, NDWI = (Green − NIR) / (Green + NIR), characterise the surface composition of each pixel along the sealed and hydrological dimensions of land cover, while the Normalised Difference Moisture Index, NDMI = (NIR − SWIR1) / (NIR + SWIR1), tracks the moisture status of vegetation and soil. Surface albedo is calculated from the surface-reflectance bands expressed as: (NIR + SWIR1) / [2 × (1 − RED)]. The value ranges from 0–1 interval. It expresses the proportion of incoming short-wave radiation that the surface returns.

Table 3.7 enumerates the nine raster layers exported at this stage of the workflow, with their value ranges and analytical roles. All are exported as GeoTIFF at 30 m resolution in EPSG:32632 — the same spatial reference and resolution as the Land Surface Temperature raster — so that every layer arrives at the 100 m study grid used for the Local Climate Zone and morphological analysis of Chapter 4 through a single, uniform resampling operation.

Layer	Range / unit	Role
NDVI	−1 to +1	Predictor; input to emissivity chain
NDBI	−1 to +1	Urban density predictor
NDWI	−1 to +1	Water-body predictor
NDMI	−1 to +1	Vegetation moisture predictor
Albedo	0 to 1	Reflectance predictor
PVI	0 to 1	Intermediate: emissivity

Emissivity (ε)	0.986–0.990	Intermediate: LST
LST	°C	Predictor target (Ch. 5); input to SUHI/SUHII
SUHII	z-score	Final SUHI intensity map

Table 3.7. Remote-sensing output layers exported from Google Earth Engine, with value range and role in the thesis.

Figure 3.6 to Figure 3.8 map the NDVI, NDBI and albedo layers listed in Table 3.7 for all four seasons in 2025. The summer NDVI raster (Figure 3.6) shows the sharpest contrast between the densely vegetated hill and river-corridor areas, where values approach 0.99, and the built-up core, where values fall to slightly negative levels; winter compresses this range and lifts NDVI over much of the built-up area towards the middle of the scale. Spring and autumn, meanwhile, both stay close to the summer pattern — a reflection of the fact that the March and October acquisition dates still fall within the growing season for most vegetation around Turin. The NDBI raster (Figure 3.7) mirrors the NDVI pattern in reverse across all four dates, with the built-up core and the industrial fringe to the north-east registering the highest values throughout the year, while the winter raster in particular is visibly noisier over the surrounding countryside, where senesced vegetation and bare soil are more easily confused with built surfaces. Surface albedo (Figure 3.8) is lowest over the dense urban fabric and the river channels in all four seasons and highest over the agricultural belt to the south-west, with the winter and spring scenes recording the strongest reflectance at these rural locations, consistent with residual low sun angle and any transient snow or bare-soil cover.

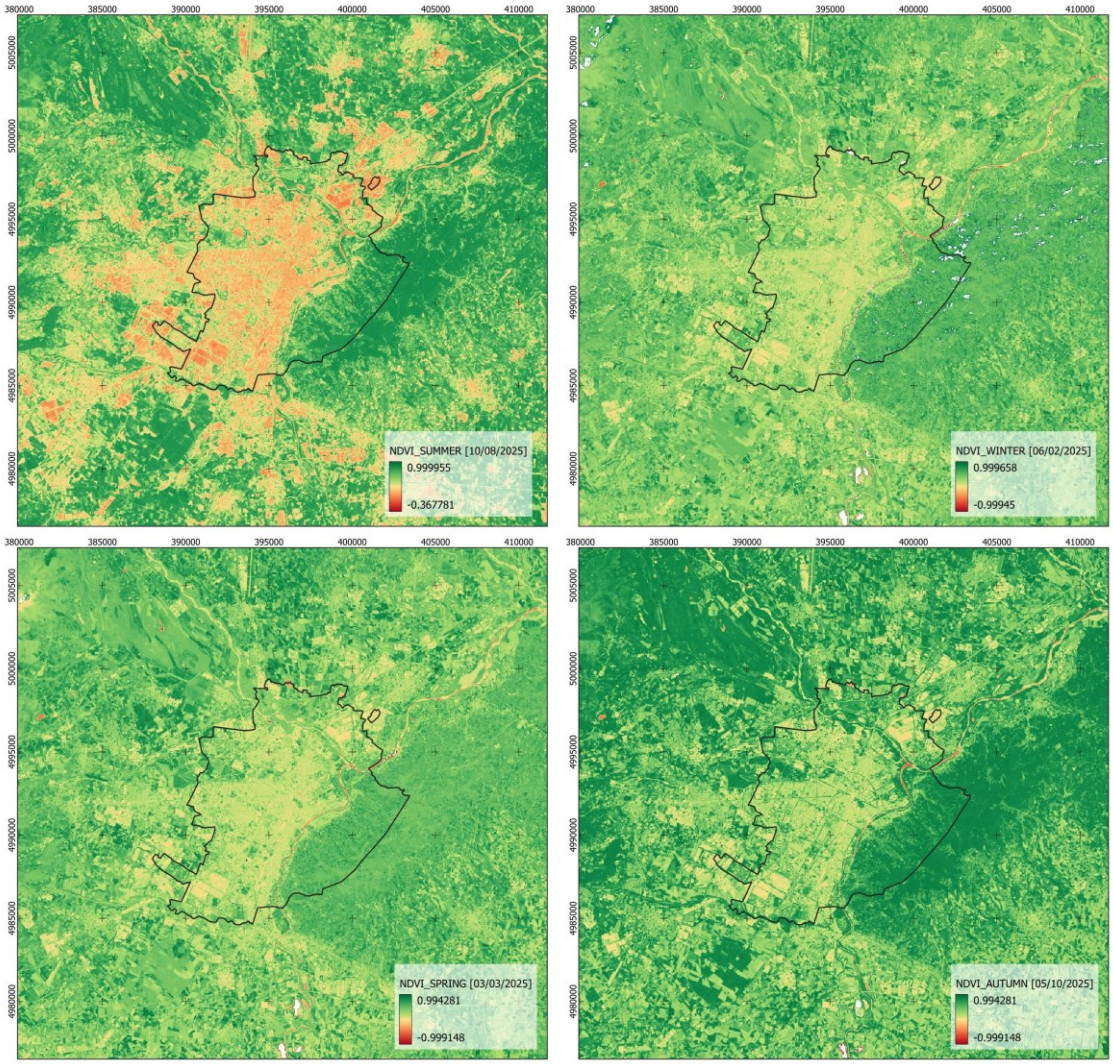


Figure 3.6. Normalised Difference Vegetation Index (NDVI) for all four seasonal 2025 scenes.

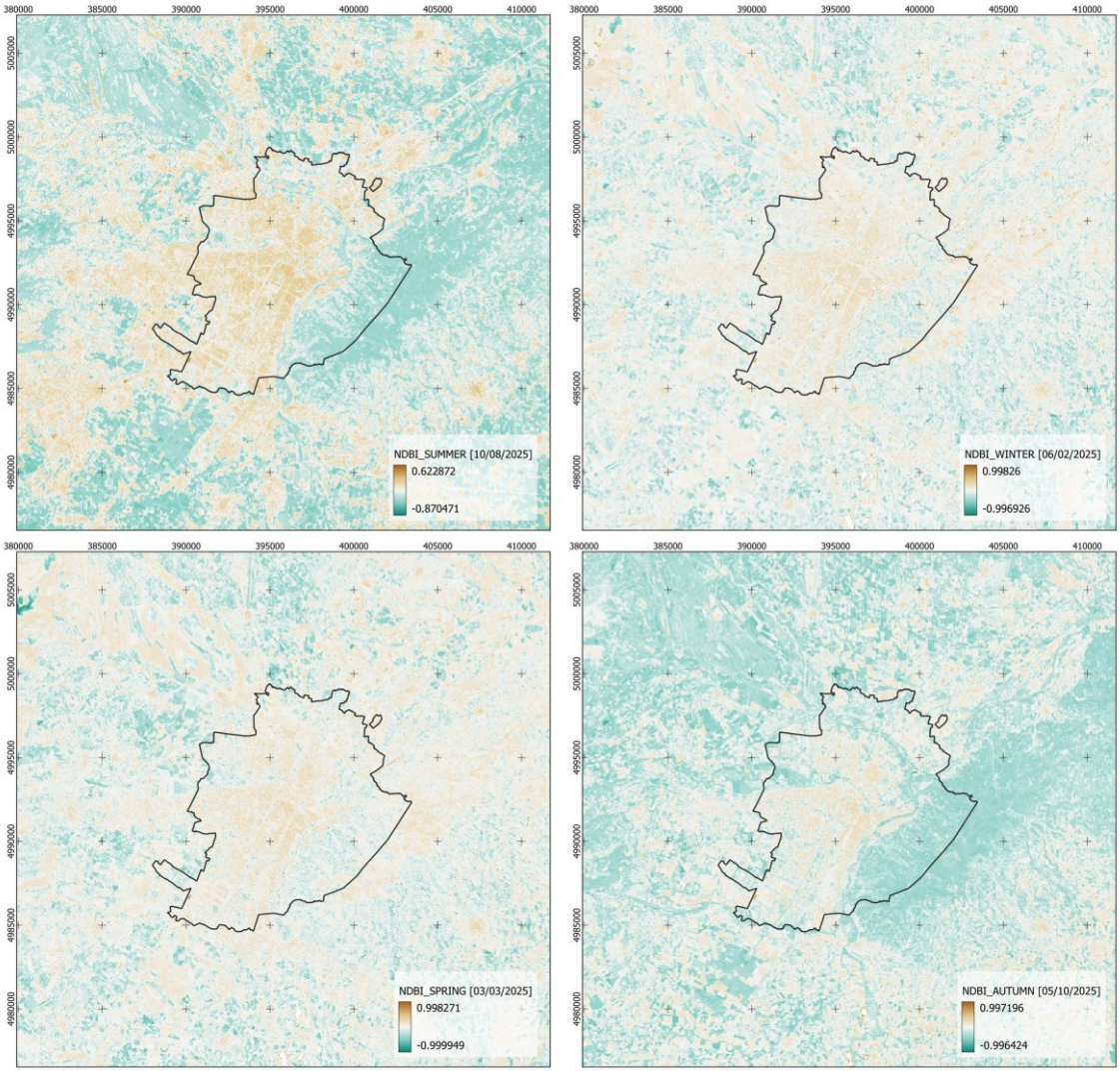


Figure 3.7. Normalised Difference Built-up Index (NDBI) for all four seasonal 2025 scenes.

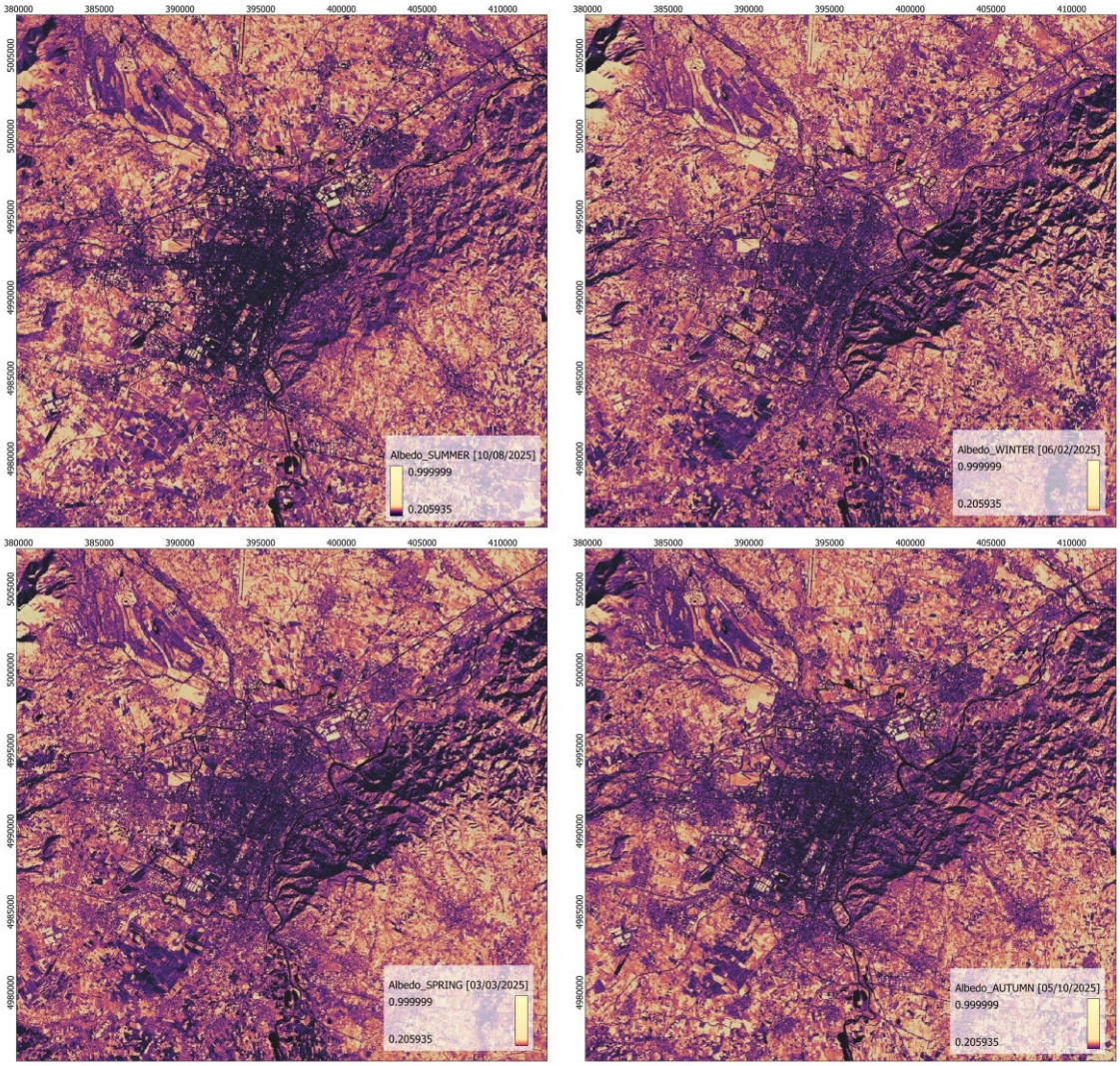


Figure 3.8. Surface albedo for all four seasonal 2025 scenes.

Figure 3.9 and Figure 3.10 complete the set of ancillary layers listed in Table 3.7, mapping NDWI and NDMI for all four seasonal 2025 scenes. The summer NDWI raster (Figure 3.9) is the clearest of the four seasons, with the Po and Dora Riparia river channels standing out as the only strongly positive values against a citywide background close to zero or negative; in winter, spring and autumn the whole area of interest trends towards higher NDWI values, consistent with wetter soil and denser low vegetation moisture outside the growing season, and the river channels become correspondingly harder to isolate visually from their surroundings. NDMI shows a related yet distinct pattern (Figure 3.10). Across

every season, it is observed that the built-up core holds the lowest and most negative values. This verifies that the impervious surfaces retain little moisture, while vegetation to the north-west and east along the hill and river corridors have the highest values. It is also observed that summer pushes this built-up and vegetation contrast to its highest point as the soil and canopy moisture deficits reach their peak. This contrast reduces again as the cooler and wetter months come in.

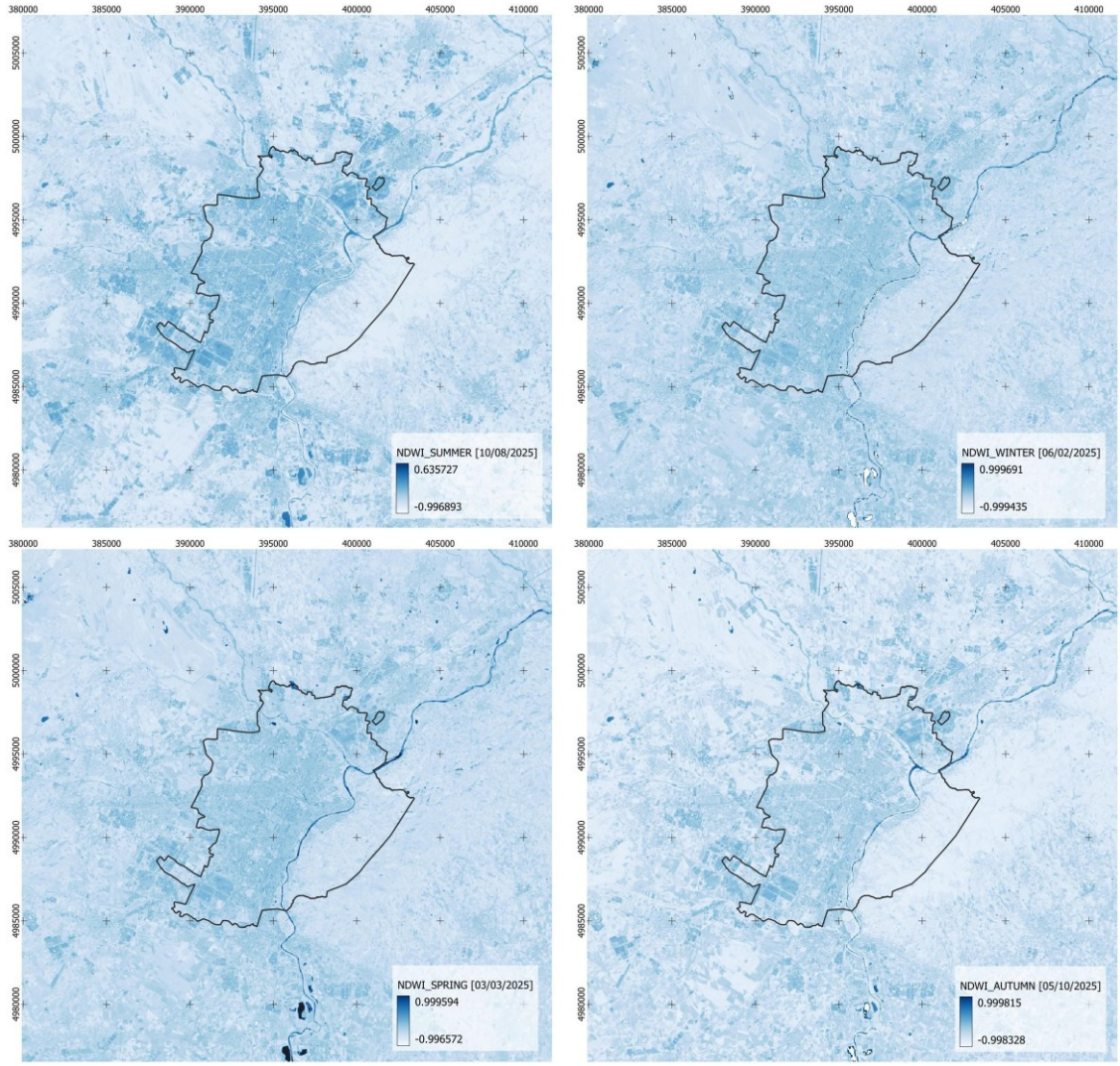


Figure 3.9. Normalised Difference Water Index (NDWI) for all four seasonal 2025 scenes.

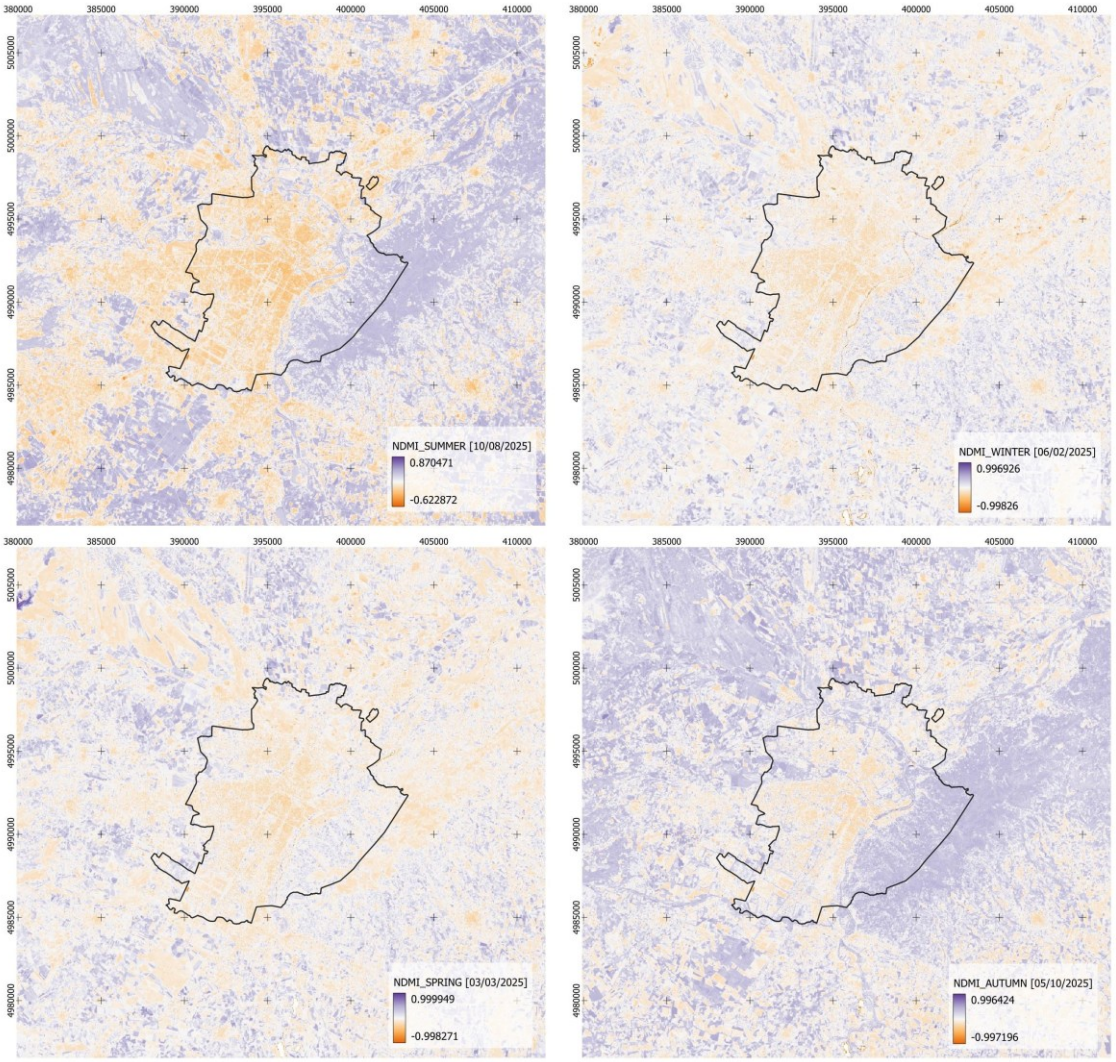


Figure 3.10. Normalised Difference Moisture Index (NDMI) for all four seasonal 2025 scenes.

3.5 Surface Urban Heat Island intensity: the SUHII z-score and the rural-reference SUHI

Two complementary measures of Surface Urban Heat Island intensity are computed from the Land Surface Temperature raster for each season. The first, the Surface Urban Heat Island Intensity (SUHII), follows the normalised z-score formulation used in the reference modelling framework for Turin:[45] SUHII = (LST − LST_mean) / LST_stdev, where the mean and standard deviation are computed

[45] (Mutani, Scalise, Sufa, & Grasso, 2024)

over the full Land Surface Temperature raster for the Turin area of interest. Being dimensionless, SUHII expresses how many standard deviations a given pixel departs from the citywide mean temperature for that season, which makes it directly comparable across seasons even though the underlying temperature range changes considerably between summer and winter.

The second measure, the absolute Surface Urban Heat Island (SUHI), is calculated in degrees Celsius relative to a fixed rural anchor: SUHI(pixel) = LST(pixel) − LST_rural, where LST_rural is the mean Land Surface Temperature within a 1 km buffer around the Moncalieri Baducchi weather station, the rural reference station used throughout this thesis. This absolute formulation is reported alongside SUHII for cross-checking, since it is more directly interpretable in physical terms even though it depends on how representative a single rural anchor point is of the wider rural area.

Four steps are followed to compute SUHI for each season. The coordinate reference system of the rural buffer and the Land Surface Temperature raster is first confirmed to match (EPSG:32632), since a mismatch at this stage would silently bias the zonal mean; zonal statistics are then run on the Land Surface Temperature raster within the 1 km buffer to obtain the rural mean, together with the standard deviation, minimum, maximum and pixel count for quality assurance; the SUHI raster is computed by raster subtraction at the native 30 m resolution; and the result is symbolised with a diverging colour ramp centred on zero, so that warmer-than-rural cells appear in warm colours and cooler-than-rural cells in cool colours across all four seasons on a consistent visual scale. Because rural Land Surface Temperature itself varies seasonally, this rural baseline is re-derived independently for each season.

Figure 3.11 shows SUHII for all four seasonal 2025 scenes, and Figure 3.12 the corresponding absolute SUHI. The summer SUHI layer displays values from approximately −10.5 °C, in the cooler river corridor and the irrigated agricultural belt south-west of the city, to approximately +23.3 °C in the most built-up sections of the urban core, a displayed range that narrows somewhat in winter (approximately −16.8 °C to +10.4 °C) as the seasonal contrast between vegetated and built-up surfaces weakens, even though the winter minimum itself falls further below the rural reference than the summer minimum does. The spring and autumn layers display values between these two extremes, at approximately −19.3 °C to +14.5 °C and approximately −22.4 °C to +12.8 °C respectively, with both transitional seasons showing a wider negative tail than either summer or winter, consistent with the still-vegetated river corridor cooling relative to a built-up core that has already begun warming. The same pattern is visible in the dimensionless SUHII maps, which display values from approximately −2.4 to +4.6 in summer, approximately −12.1 to +8.1 in winter, approximately −9.8 to +7.1 in spring and approximately −7.8 to +5.8 in autumn; the Po and Dora Riparia corridors register the most negative displayed z-scores in summer and the built-up core and its north-eastern fringe the most positive ones, consistent with the physical SUHI values reported above.

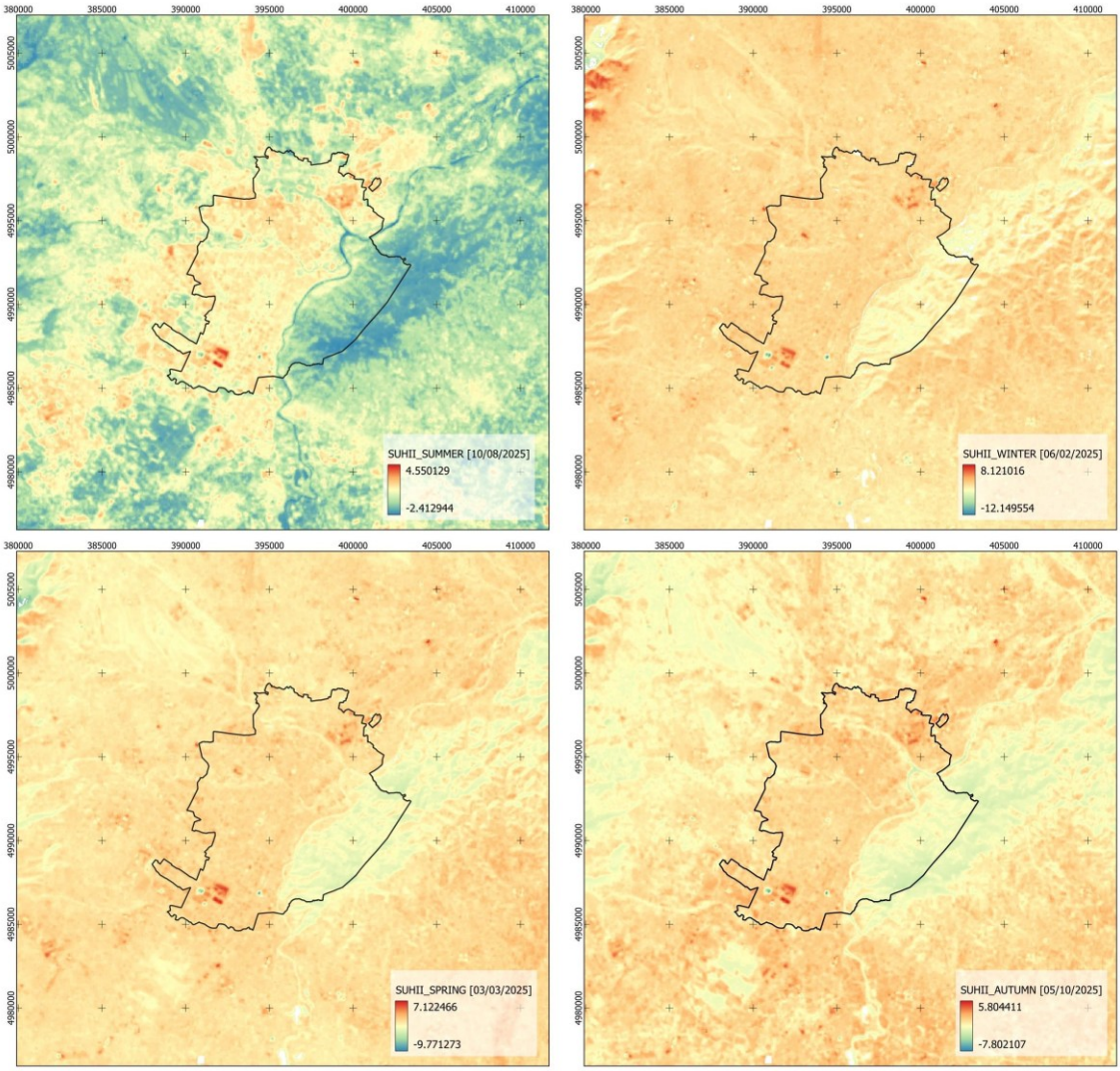


Figure 3.11. SUHII (z-score) for all four seasonal 2025 scenes.

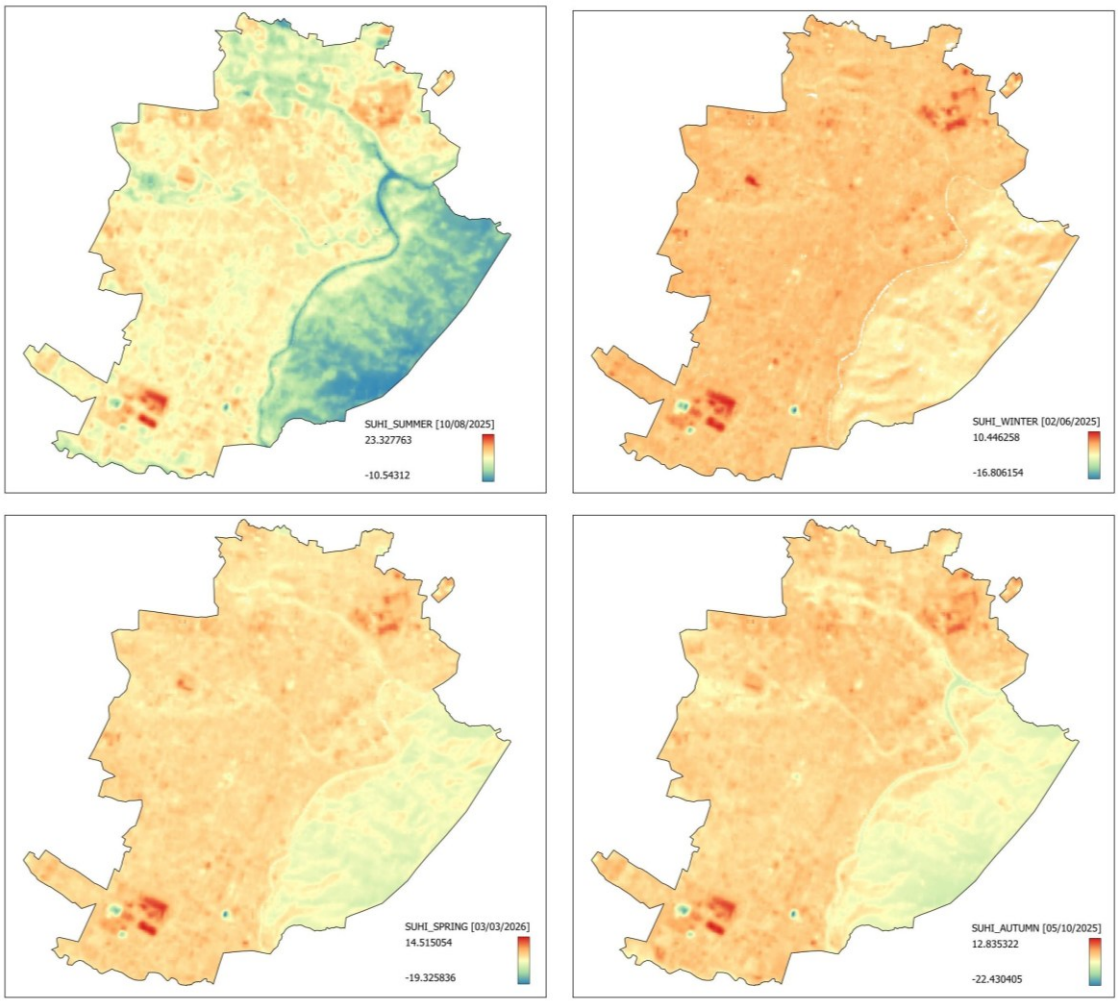


Figure 3.12. Absolute, rural-reference SUHI (°C) for all four seasonal 2025 scenes, relative to the Moncalieri Baducchi station.

3.6 Urban Thermal Field Variance Index

A complementary ecological perspective on the same Land Surface Temperature raster is provided by the Urban Thermal Field Variance Index (UTFVI), proposed by Liu and Zhang[46] as a way of expressing each pixel's thermal condition relative to the mean of the whole study area: UTFVI = (LST − LST_mean) / LST_mean. SUHI and SUHII are primarily descriptive of the urban–rural temperature contrast, whereas UTFVI is conventionally classified into six ecological-stress classes — here labelled Best, Better, Normal, Bad, Worse and

[46] (Liu & Zhang, 2011)

Worst — that translate the continuous index into a form more directly usable for identifying where heat is likely to pose the greatest ecological and human-comfort burden, following the numeric thresholds set out below.

Figure 3.13 shows the classified UTFVI raster for all four seasonal 2025 scenes. In summer, the Bad to Worst classes cover much of the built-up core and extend along the north-eastern industrial fringe, broadly overlapping the areas already identified as having the highest SUHI and SUHII values. In winter, the great majority of the area of interest falls into the Best or Better classes instead, with only a small number of localised hotspots, concentrated in the same built-up sections that stand out in summer, remaining in the higher-stress classes. Spring and autumn each retain a visibly larger higher-stress footprint than winter without reaching the summer extent, so that the same recurring cluster of built-up locations moves gradually in and out of the Bad-to-Worst classes across the four dates, instead of a simple two-state switch between summer and winter alone. This seasonal persistence of a recurring set of hotspots is examined more formally below and is used in Chapter 6 to help prioritise where LCZ-specific mitigation measures are targeted first.

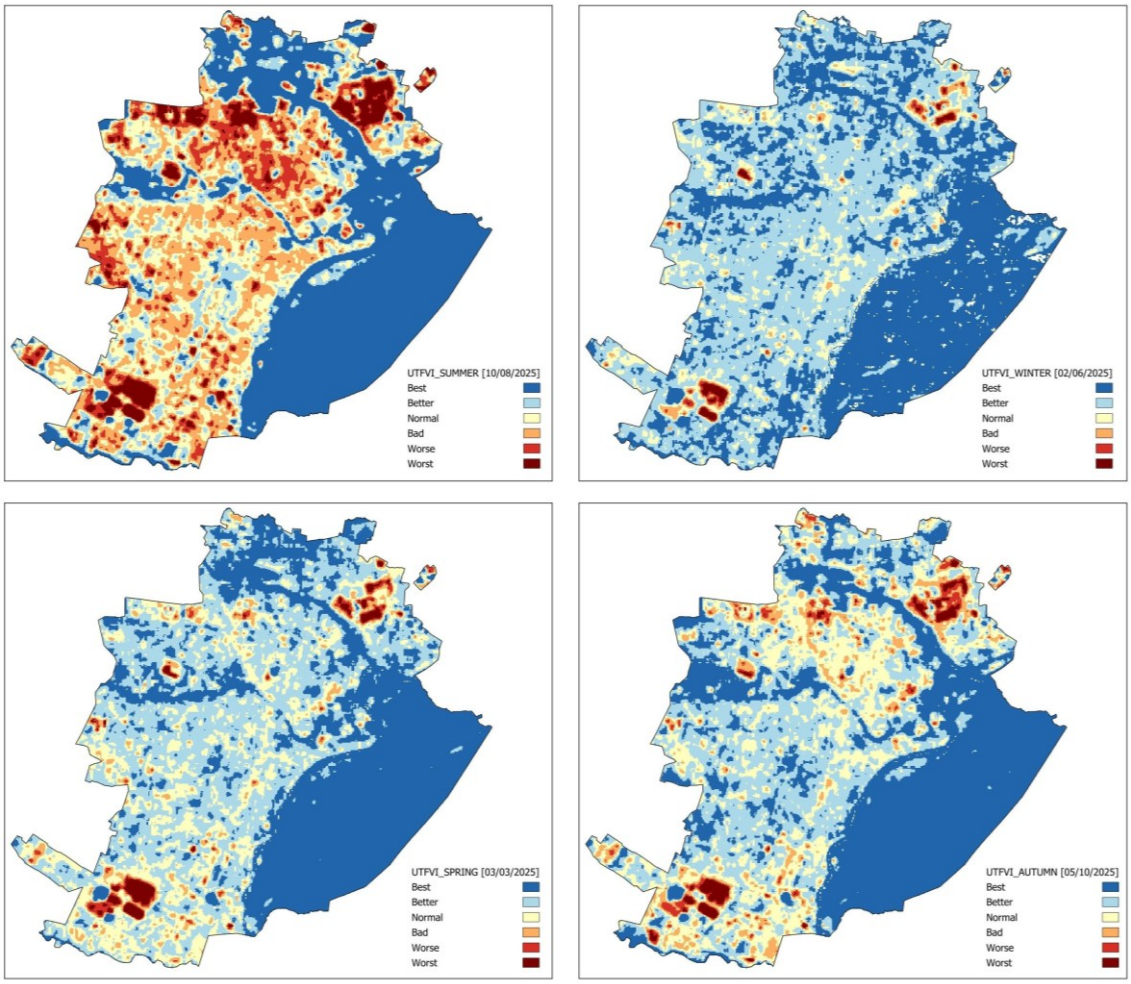


Figure 3.13. Urban Thermal Field Variance Index, classified into six ecological-stress levels, for all four seasonal 2025 scenes.

UTFVI characterises each pixel's thermal severity on a standardised ecological scale, though it does so relative to the intra-urban temperature distribution rather than to the rural reference condition. Thermal hotspots are therefore identified in QGIS by reading the SUHI and UTFVI layers for each seasonal Landsat scene in combination. A pixel qualifies as a hotspot when the following conditions are jointly satisfied: a SUHI value > zero, placing the surface above the rural reference temperature, and a UTFVI value at or above the threshold that describes the ecologically stressed end of Zhang, Wang and Li's six-tier scheme.

So, the UTFVI classification falling within the Bad, Worse, or Worst classes [47] are considered as Hotspots.

Therefore, Hotspot can be expressed as: Hotspot = 1 if (SUHI > 0) AND (UTFVI ≥ 0.010), otherwise Hotspot = 0.

UTFVI is computed as (LST − LST_mean) / LST_mean, with units in kelvin. The value 0.010 represents the lower boundary of the Bad class, within the full six-tier gradation. Excellent means below 0, Good is 0 to 0.005, Normal is classified as values from 0.005 to 0.010, Bad is 0.010 to 0.015 class, Worse is 0.015 to 0.020, and Worst is above 0.020. Each of the four seasonal scenes that is summer, autumn, winter, spring goes through this procedure separately, producing four hotspot layers.

Figure 3.14 maps how these hotspots persist across the year. Adding the four seasonal hotspot layers together gives a single composite annual raster output which is the annual Hotspot. Hotspot_annual = Hotspot_summer + Hotspot_autumn + Hotspot_winter + Hotspot_spring, with values ranging from 0 to 4 depending on number of seasons a given pixel qualifies as hotspot. 0 means it never qualifies as hot spot whereas 4 means it does in every season. The persistent class concentrates in the same built-up sections of the urban core and along the north-eastern industrial fringe that already stand out in the summer UTFVI and SUHI rasters, together with a cluster to the south-west of the city that recurs across the seasonal SUHII and SUHI maps shown earlier in this chapter; this measure of seasonal persistence is carried forward to support the spatial-priority interpretation developed in Chapter 6.

[47] (Zhang, Wang, & Li, 2006)

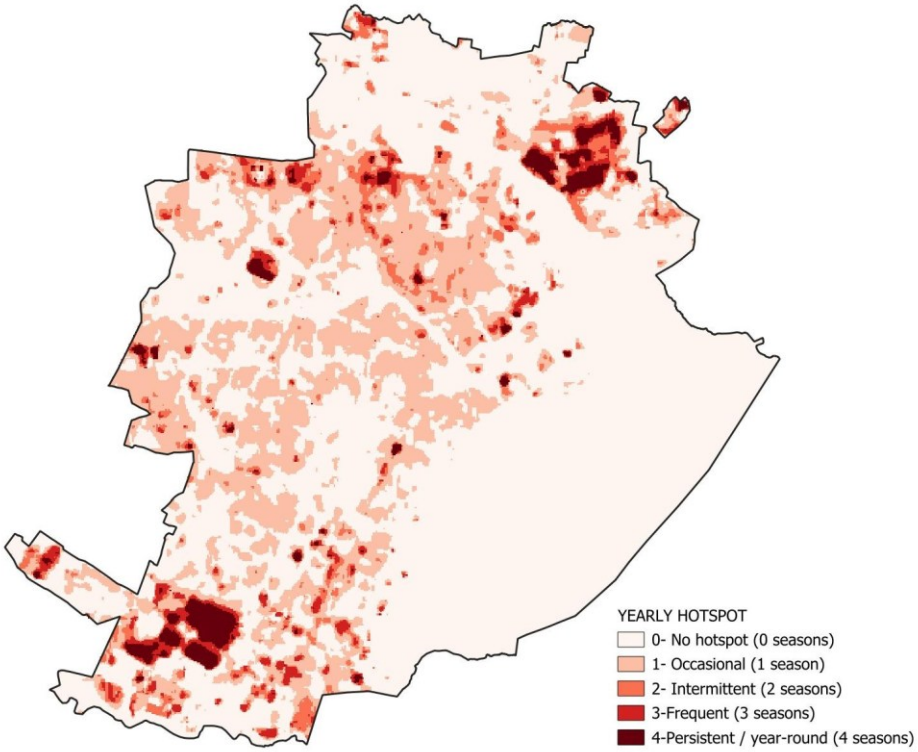


Figure 3.14. Yearly persistence of SUHI/UTFVI-based thermal hotspots across the four seasonal scenes, classified from never a hotspot (0 seasons) to a hotspot in every season (4 seasons).

The Local Climate Zone classification and the locally derived morphological indicators against which these thermal patterns are interpreted in this thesis are introduced in Chapter 4, together with the morphological indicators computed independently from Turin-specific building and land-cover data.

Chapter 4- Local Climate Zones and Urban Morphology of Turin

4.1 The Demuzere et al. (2022) global LCZ classification and its adaptation to Turin

The Local Climate Zone scheme classifies urban and rural landscapes into seventeen standard classes, ten of them built types distinguished primarily by building height and packing density and seven of them land-cover types distinguished by vegetation, terrain and water.[48] Each class is defined by measurable ranges of surface-cover and surface-structure properties, including sky view factor, aspect ratio, building surface fraction, impervious surface fraction, pervious surface fraction and the height of roughness elements, so that a given class carries a consistent physical meaning wherever it is applied. This thesis uses the global Local Climate Zone classification of Demuzere, Kittner and Bechtel,[49] an extension of the earlier European-scale mapping exercise of Demuzere et al.,[50] which applies a Random Forest classifier trained on expert-labelled training areas to Sentinel and Landsat imagery to produce a worldwide, 100 m resolution Local Climate Zone map at native 30 m classification resolution. The specific layer used in this thesis is the filtered layer distributed as lcz_filter_v3.tif, which applies a morphological Gaussian filter to the raw per-pixel classification. This filter uses distance-weighted voting, in which nearer pixels carry more influence than more distant ones under a bell-shaped kernel, and is class-aware in the sense that the effective search radius accounts for the typical

[48] (Stewart & Oke, 2012)

[49] (Demuzere, Kittner, & Bechtel, A global map of local climate zones to support earth system modelling and urban-scale environmental science, 2022)

[50] (Demuzere, Bechtel, Middel, & Mills, 2020)

patch size of each Local Climate Zone type. The filtered layer better preserves linear features such as rivers and roads than a simple majority filter would, and its effective resolution of approximately 85–100 m is directly compatible with the 100 m study grid used throughout this thesis. The raster arrives in geographic coordinates (EPSG:4326) and gets reprojected to EPSG:32632 (WGS 84 / UTM Zone 32N), bringing it into alignment with both the Landsat-derived indicators from Chapter 3 and the locally derived morphological indicators covered in Section 4.2.

The Local Climate Zone raster is classified in integers from 1-17. Out of which 1–10 covers the ten built types and 11–17 (also expressed as A–G) cover the seven land-cover types. Example Class 14 corresponds to LCZ D (low plants) in Stewart and Oke's original nomenclature. This identification is followed consistently throughout the thesis. Figure 4.1 shows the classified Local Climate Zone raster clipped to Turin's area of interest, alongside the Moncalieri Baducchi weather station and its 1 km buffer, used as the rural reference in Chapter 3. Classes 2 and 3 which are compact midrise and compact low-rise dominate Turin's built-up core, while class 5 (open midrise) and class 6 (open low-rise) take over as the residential fringe spreads outward. Along the Po and Dora Riparia river corridors, and across the agricultural belt to the south-west and north, class 14 (LCZ D, low plants) is the predominant classification — this is also where the Moncalieri Baducchi weather station sits.

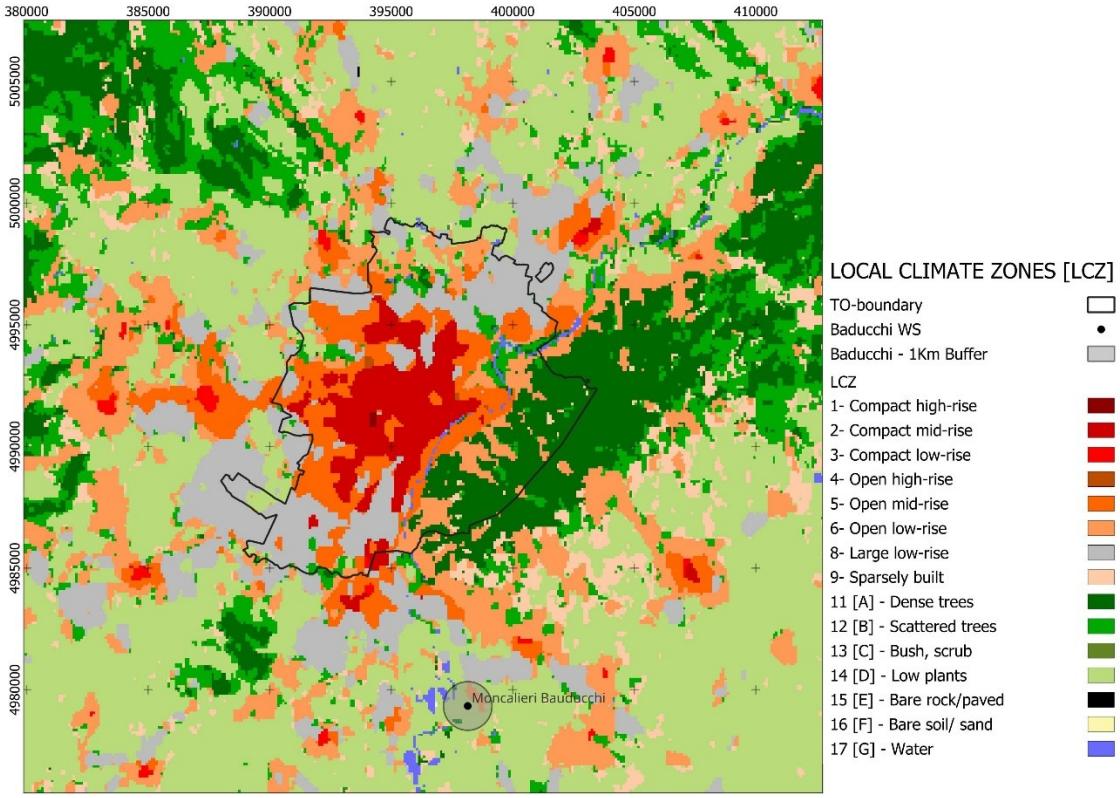


Figure 4.1. Local Climate Zone classification for the Turin area of interest (Demuzere et al., 2022), showing the Moncalieri Baducchi rural reference station and its 1 km buffer.

4.2 Locally derived morphological indicators

Because a global Local Climate Zone classification necessarily generalises local building form, this thesis also derives a parallel set of morphological indicators directly from Turin-specific building and land-cover data, following the general approach used in earlier urban-scale studies of the city.[51] Building footprints, eave heights and ground elevations are taken from the Carta Tecnica Regionale via the Geoportale Torino, from which the height (H_ed), footprint area (A_ed) and volume (V_ed) of each building are calculated. These building-level values are then aggregated to a 100 m grid to obtain the Building Coverage Ratio (BCR), Green Coverage Ratio (GCR), Impervious Surface

[51] (Mutani, Cristino, & Bullita, The Urban Heat Island of the Metropolitan City of Turin: Strategies for a Sustainable Urban Planning, 2018)

Fraction (ISF) and Building Density (BD), and a 100 m distance matrix, restricted to the thirty nearest neighbours within 100 m of each cell, is used to derive the relative building height (H/Havg), the mean street or building spacing (W) and the aspect ratio (H/W). Table 4.1 sets out the formula used for each indicator.

Variable	**Formula**
H_ed (building height)	QT_GRONDA − QT_SUOLO
A_ed (footprint area)	area($geometry)
V_ed (building volume)	H_ed × A_ed
BCR (building coverage ratio)	Sum(A_ed) / 10,000
GCR (green coverage ratio)	Sum (veg. area) / 10,000
ISF (impervious surface fraction)	Sum (impervious area) / 10,000
BD (building density)	Sum(V_ed) / 10,000
BH (volume-weighted building height)	Sum (H_ed × V_ed) / Sum(V_ed)
SVF (sky view factor)	Derived from a DSM using the SVF calculator in the UMEP QGIS plugin, buildings only
H/Havg (relative building height)	Sum [(H_ed / mean H_ed) × V_ed] / Sum(V_ed)
W (mean spacing)	Mean distance to the 30 nearest neighbours within 100 m
H/W (aspect ratio)	Sum [(mean H / W_ed) × V_ed] / Sum(V_ed)

Table 4.1. Morphological Indicators derived from Geoportale Torino data on a 100 m grid.

Figure 4.2 shows the resulting Sky View Factor raster: the lowest values, indicating the most enclosed street canyons, coincide closely with the compact built-up core already identified as classes 2 and 3 in the Demuzere Local Climate Zone map, while the highest values, close to 1, are recorded over the open agricultural land to the south-west and over the largest parks. Figure 4.3 shows the Building Coverage Ratio and Building Density maps, which display a broadly

concentric pattern of decreasing built-up intensity away from the historic centre, interrupted by secondary industrial and commercial clusters in the northern and south-western parts of the area of interest. Figure 4.4 shows the corresponding Aspect Ratio and relative building height maps, in which the highest H/W and H/Havg values are concentrated along the same dense central streets that register the lowest Sky View Factor, consistent with the expectation that narrower, more enclosed canyons correspond to a higher ratio of building height to street width.

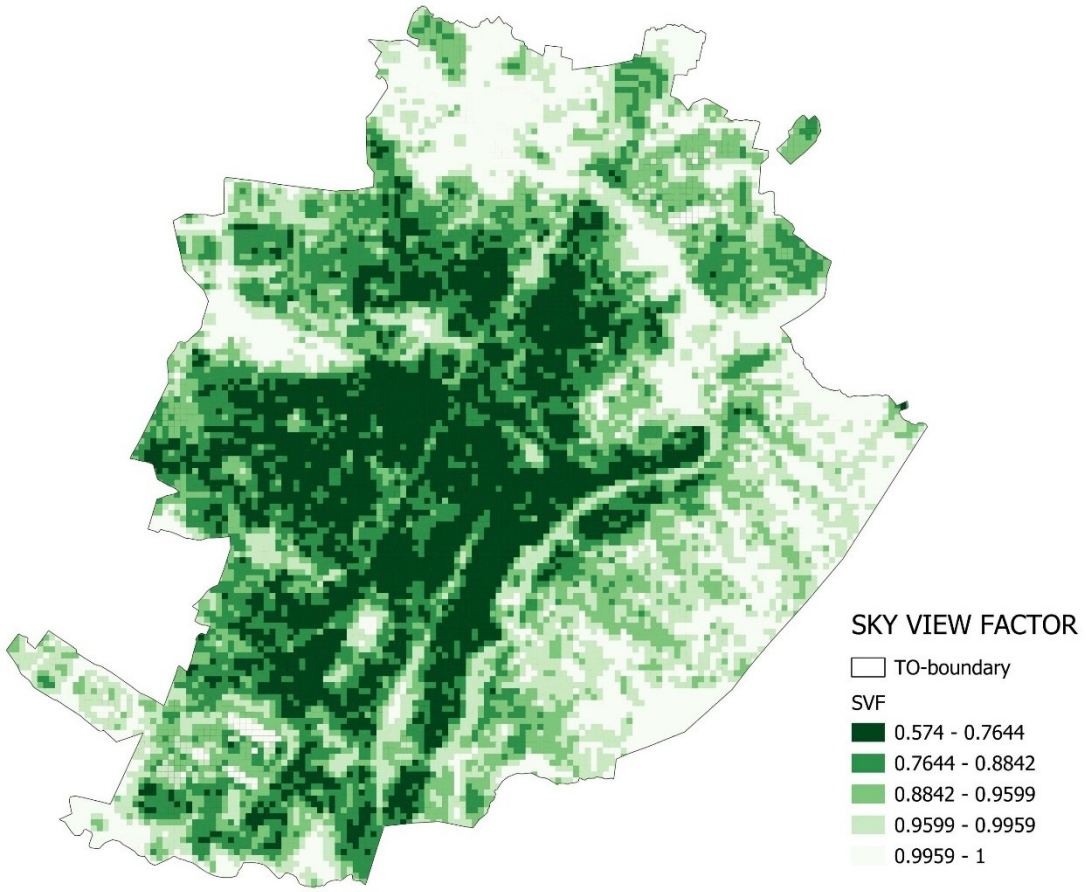


Figure 4.2. Sky View Factor derived from the Turin digital surface model (buildings only).

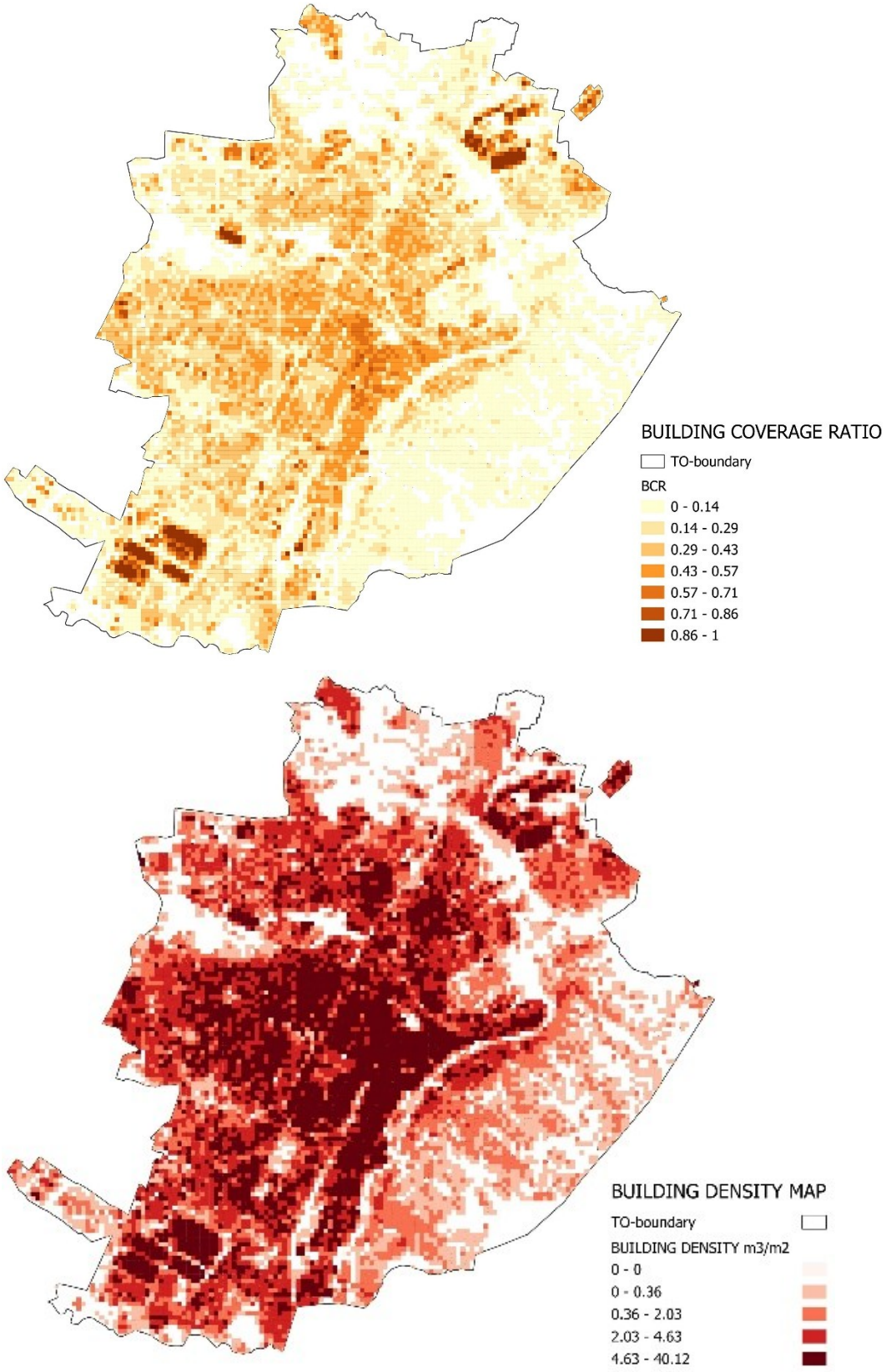


Figure 4.3. Building Coverage Ratio and Building Density- aggregated to the 100 m grid.

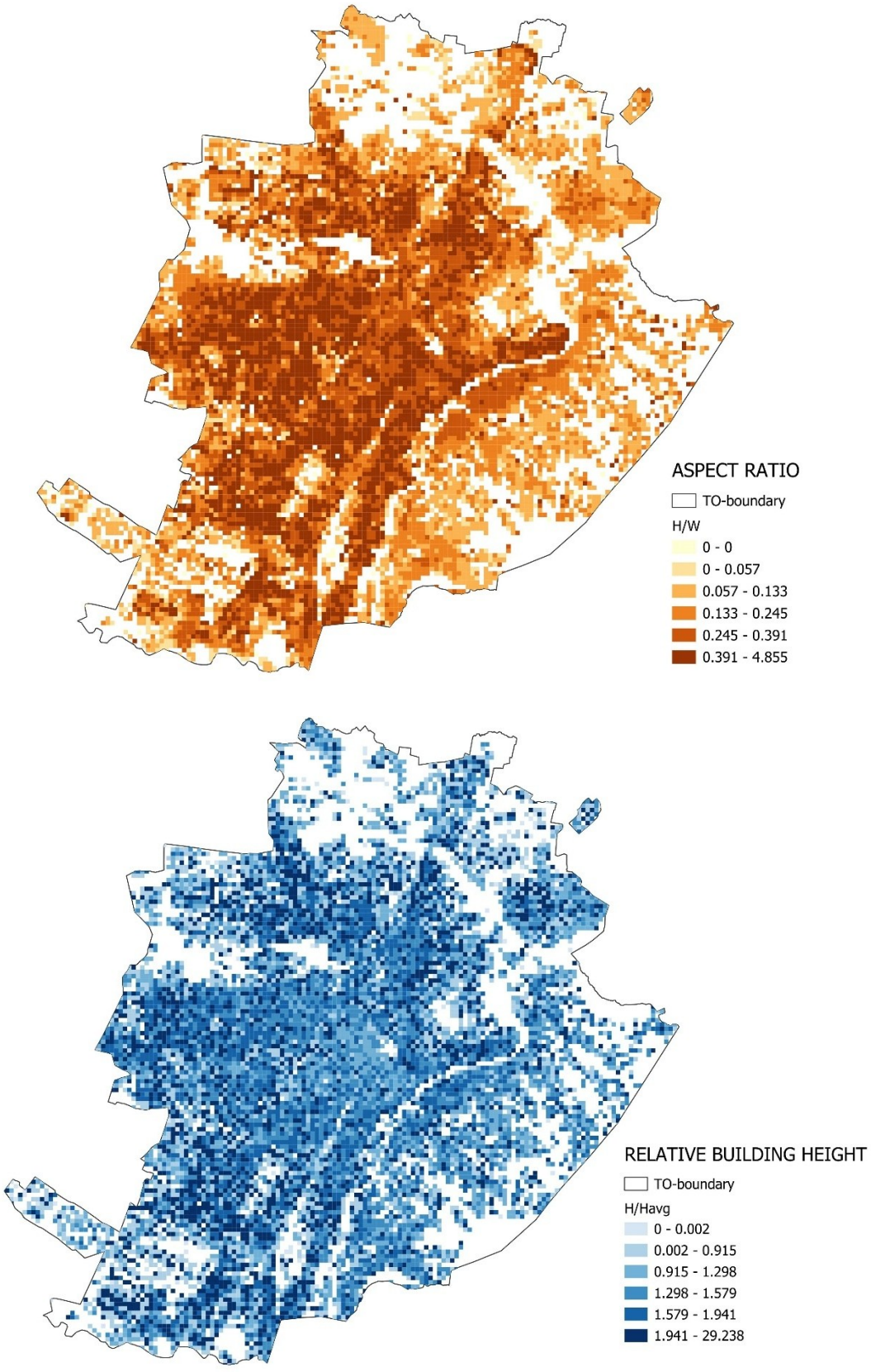


Figure 4.4. Aspect ratio (H/W) and relative building height (H/Havg), derived from the 100 m distance matrix.

All locally derived indicators are finally resampled onto the same 100 m study grid used for the Land Surface Temperature, SUHI, SUHII and UTFVI rasters of Chapter 3, so that every predictor available for the Random Forest modelling of Chapter 5 shares a common spatial reference without a further reprojection step.

4.3 Relationship between the global LCZ classification and locally derived morphology

A global Local Climate Zone classification trades local precision for consistent, worldwide coverage. A directly comparable precedent for checking such a classification against locally derived morphological data is provided by a comparison of the WUDAPT Level-0 Local Climate Zone classification against a morphology-based Local Climate Zone classification for several Italian cities, including Lecce, Bari, Naples, Rome and Milan.[52] That comparison found that the two approaches agreed reasonably well on the overall spatial pattern of classes, but that the WUDAPT classification tended to misclassify some cells at the dense urban core, while wider agreement was observed at the transition between urban and rural areas; the morphology-based classification, in turn, required a level of detailed elevation and building data that is not available for every city.

For Turin, this thesis does not repeat that exercise as a formal validation step: the Demuzere Local Climate Zone classification is instead used directly, as distributed, for the Random Forest modelling of Chapter 5, and the locally derived morphological indicators of Section 4.2 are treated as an independent predictor set, and are not used to correct or re-classify the global classification.

[52] (Buccolieri, et al., 2022)

Table 4.2 reproduces the property ranges that Stewart and Oke[53] associate with each Local Climate Zone class present in the Turin area of interest, and is retained here as reference material: it clarifies what physical conditions a given Demuzere-assigned class is expected to represent, which is useful when interpreting the results of Chapters 5 and 6.

LCZ class	Sky view factor	Aspect ratio	Building surface fraction (%)	Impervious surface fraction (%)	Pervious surface fraction (%)	Height of roughness elements (m)
2 – Compact midrise	0.3–0.6	0.75–2	40–70	30–50	<20	10–25
3 – Compact low-rise	0.2–0.6	0.75–1.5	40–70	20–50	<30	3–10
5 – Open midrise	0.5–0.8	0.3–0.75	20–40	30–50	20–40	10–25
6 – Open low-rise	0.6–0.9	0.3–0.75	20–40	20–50	30–60	3–10
8 – Large low-rise	>0.7	0.1–0.3	30–50	40–50	<20	3–10
9 – Sparsely built	>0.8	0.1–0.25	10–20	<20	60–80	3–10
11 (A) – Dense trees	<0.4	>1	<10	<10	>90	3–30
12 (B) – Scattered trees	0.5–0.8	0.25–0.75	<10	<10	>90	3–15
13 (C) – Bush, scrub	0.7–0.9	0.25–1.0	<10	<10	>90	<2
14 (D) – Low plants	>0.9	<0.1	<10	<10	>90	<1

Table 4.2. Stewart and Oke (2012) property ranges for the Local Climate Zone classes present in the Turin area of interest. All properties are unitless except the height of roughness elements (m).

A quantitative, cell-by-cell comparison of this kind, following the approach of Buccolieri et al.,[54] would be a natural extension of this chapter; it has not been

[53] (Stewart & Oke, 2012)

[54] (Buccolieri, et al., 2022)

carried out for this thesis; and it is noted as a direction for future refinement in Chapter 7.

4.4 The rural reference station within the LCZ scheme

The Local Climate Zone classification also provides an independent check on the choice of rural and urban reference stations used throughout this thesis. The Moncalieri Baducchi station, used as the fixed rural anchor for the absolute SUHI calculation in Section 3.5, falls within class 14 (LCZ D, low plants) in the Demuzere classification, consistent with the low building coverage, high Sky View Factor and predominantly vegetated surroundings expected of that class and with its established role as the rural reference station for this thesis.

The Via della Consolata station, introduced in Chapter 2 as the urban climate reference for Turin, by contrast falls within the compact built-up classes at the historic centre of the city, consistent with its location in a dense, low-rise to midrise urban canyon. Read together with the locally derived morphological indicators of Section 4.2, this confirms that the two reference stations used throughout this thesis sit at the extremes of the Local Climate Zone distribution present in the Turin area of interest, which is a desirable property for a study that aims to relate Surface Urban Heat Island intensity to the full range of urban morphology captured by the Local Climate Zone scheme.

The combined set of predictors introduced in this chapter and in Chapter 3 — the remote-sensing indicators, the Local Climate Zone class and the locally derived morphological indicators — forms the input to the Random Forest modelling described in Chapter 5, where their relative importance in explaining the spatial and seasonal pattern of SUHI and SUHII is assessed using Pearson correlation, variance inflation factor screening and SHAP interpretation.

Chapter 5- Modelling the Drivers of Surface Urban Heat Island Intensity

5.1 Purpose and Analytical Approach

Chapter 4 established the morphological and spectral characterisation of Turin's local climate zones, giving a descriptive account of how building form, surface cover and land use vary across the study area. This chapter turns that description into a predictive one, investigating the relationships between urban morphology, climate and surface characteristics on one hand and the surface urban heat island on the other. Two related aims follow from this. The first is to identify which variables contribute most strongly to the spatial differences in surface urban heat island intensity observed across Turin. The second is to do so using an interpretable form of machine learning, so that the resulting model is predictive, and can also be examined to understand how each predictor shapes its output.

Random Forest regression, interpreted through Shapley Additive exPlanations (SHAP), is used throughout this chapter to meet both aims. Breiman established that Random Forest captures non-linear and interaction effects among morphological and spectral predictors without needing a parametric functional form specified beforehand, therefore it was selected here [55]. Building on Lundberg and Lee's work, [56] SHAP breaks each individual predictor down into additive per-predictor contributions, grounded in cooperative game theory. Gini-based feature importance, taken from the same fitted trees gives a single average score for each predictor's contribution. The chapter follows the

[55] (Breiman, 2001)

[56] (Lundberg & Lee, 2017)

sequence in which the work was carried out: Section 5.2 describes the modelling dataset and its spatial framework; Section 5.3 introduces the candidate predictor variables; Section 5.4 documents the multicollinearity screening that reduced the candidate set; Sections 5.5 to 5.7 present the season-specific Random Forest models, their performance and their interpretation; Section 5.8 synthesises the results across the four seasons; and Section 5.9 draws out the implications for the mitigation-scenario work of Chapter 6, without developing that methodology here.

5.2 Modelling Dataset and Spatial Framework

The modelling dataset was assembled by joining, at the level of each 100 × 100 m grid cell, the spectral indicators derived in Chapter 3 with the morphological indicators derived in Chapter 4 from the Demuzere LCZ classification and the Geoportale Torino building and land-cover layers. The resulting grid covers the Turin metropolitan area with approximately 13,447 cells, a figure that varies marginally by season (13,444 to 13,447 cells) after cloud and boundary masking. Four seasonal datasets, one for each of the representative dates fixed in Chapter 3 for 2025, the analysed representative year in the current workflow, are modelled independently throughout this chapter, so that seasonal differences in the drivers of surface heat can be examined directly, without being averaged away.

Following the definitions given in Chapter 3, every grid carries two response variables SUHI and SUHII. SUHI, expressed in degree Celsius, captures the absolute surface urban heat island which is given by the difference in land surface temperature between a given cell and a fixed rural reference. While SUHII gives the normalised counterpart- a z-score computed against the

seasonal mean and the standard deviation of LST across the study area. This chapter keeps the SUHI and SUHII separate throughout as SUHI holds onto physical units and magnitude needed to read an absolute temperature deviation, and SUHII puts that same signal on a scale that works across all seasons even if they differ in overall thermal contrast. Each grid cell also carries the predictor variables from Section 5.3 that are obtained from the same seasonal Landsat data and the morphological database explained in Chapter 4.

Figure 5.1 summarises the end-to-end modelling workflow followed throughout this chapter, from the data preparation stage through hyperparameter tuning, model fitting and interpretation, to the outputs carried forward into Chapter 6. The workflow is applied independently to each of the four seasons, with the predictor-screening and hyperparameter-tuning stages repeated separately each time so that every model reported in this chapter remains directly comparable with the others.

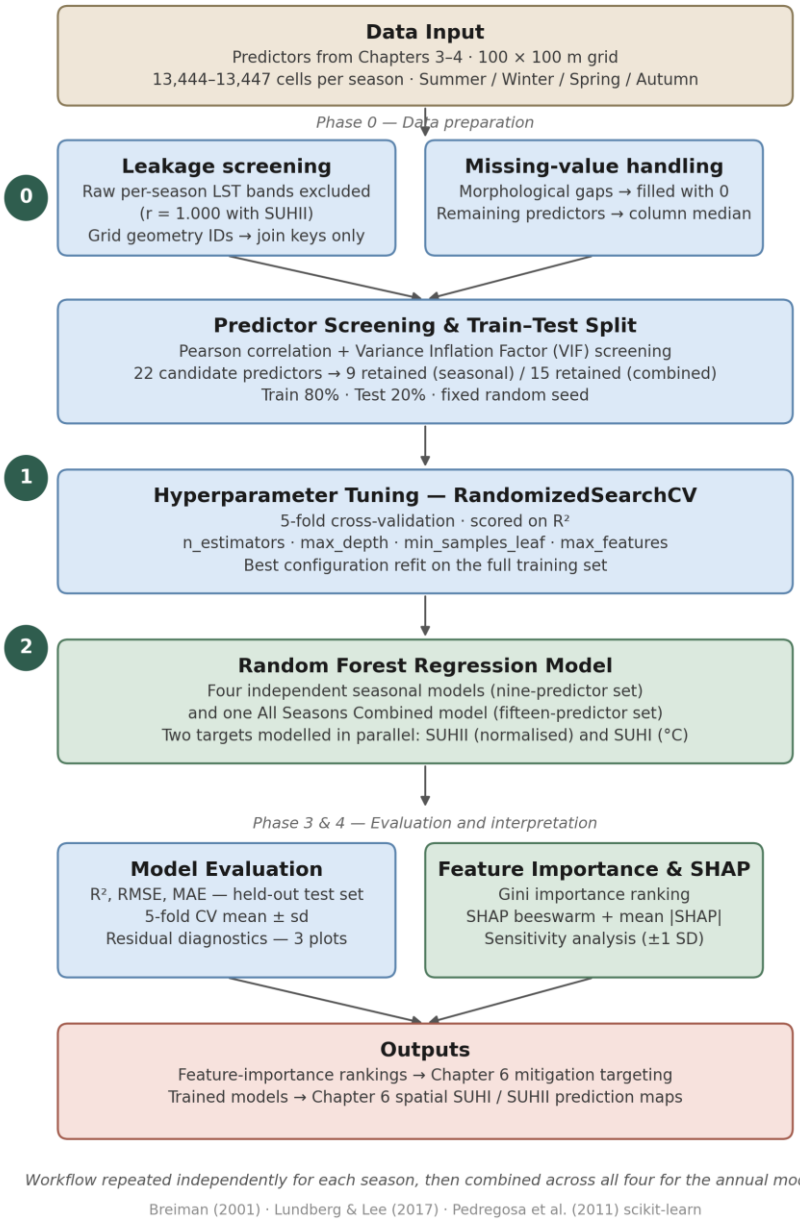


Figure 5.1. End-to-end Random Forest modelling workflow, from data preparation through hyperparameter tuning, evaluation and interpretation, to the outputs carried into Chapter 6.

5.3 Predictor Variables

Twenty-two candidate predictors entered the initial screening, grouped into three conceptual families matching the structure of Chapters 3 and 4. Nine morphological predictors describe the built form of each cell: building density, building coverage ratio, relative building height, height-to-width ratio, sky view

factor, green coverage ratio, impervious surface fraction, pervious surface fraction and population density. Six climate predictors describe the ambient meteorological conditions for the season: air temperature, relative humidity, wind speed, wind direction (resolved into sine and cosine components to preserve its circular character) and incoming solar radiation, each drawn from the single representative station value adopted for that season in Chapter 3 and broadcast uniformly across the grid. Seven spectral predictors describe the surface cover and radiative properties visible from the Landsat imagery: NDBI, NDMI, NDVI, NDWI, PVI, albedo and emissivity.

Because the climate predictors are broadcast from a single station value, they carry no spatial variance within any one season: every cell in a given seasonal dataset receives the same air temperature, the same relative humidity, and so on. This is stated explicitly here because it shapes how the climate predictors were treated at the screening stage: a Random Forest cannot use a constant column to explain why one cell is hotter than its neighbour, so the six climate predictors were confirmed to contribute negligible importance to every season-specific model, and are excluded entirely from the final predictor set on this evidence, tightening the model to the predictors that carry genuine explanatory information for a single season.

The twenty-two candidate predictors were screened for redundancy and multicollinearity before any Random Forest model was fitted, following the procedure set out in Section 5.4. Thirteen variables were removed on this evidence: PVI, emissivity, pervious surface fraction, NDMI, NDWI and green coverage ratio, together with NDBI, whose variance inflation factor was found to be infinite once the other candidate predictors were included in the same design matrix, and all six climate predictors, confirmed to carry zero within-

season spatial variance. The screening reduced the candidate set from twenty-two predictors to a final set of nine: seven morphological predictors (building density, building coverage ratio, relative building height, height-to-width ratio, sky view factor, impervious surface fraction and population density) and two spectral predictors (NDVI and albedo). This nine-predictor set is used for every Random Forest model reported in this chapter.

5.4 Multicollinearity and VIF Analysis

Multicollinearity among predictors matters here precisely because this model is meant to be interpreted, and interpretation demands more than raw predictive accuracy. When two predictors move together almost perfectly, a fitted model can assign importance to either one more or less arbitrarily, leaving the resulting ranking sensitive to which member of the pair happens to be retained. For this reason, the predictor set was screened in two complementary ways before any Random Forest model was fitted, using the twenty-two-predictor, all-seasons-combined dataset assembled for this purpose. Pairwise Pearson correlation identified predictor pairs that move together almost perfectly, while the Variance Inflation Factor (VIF), computed on an intercept-corrected design matrix, identified multivariate redundancy that a pairwise check alone can miss.

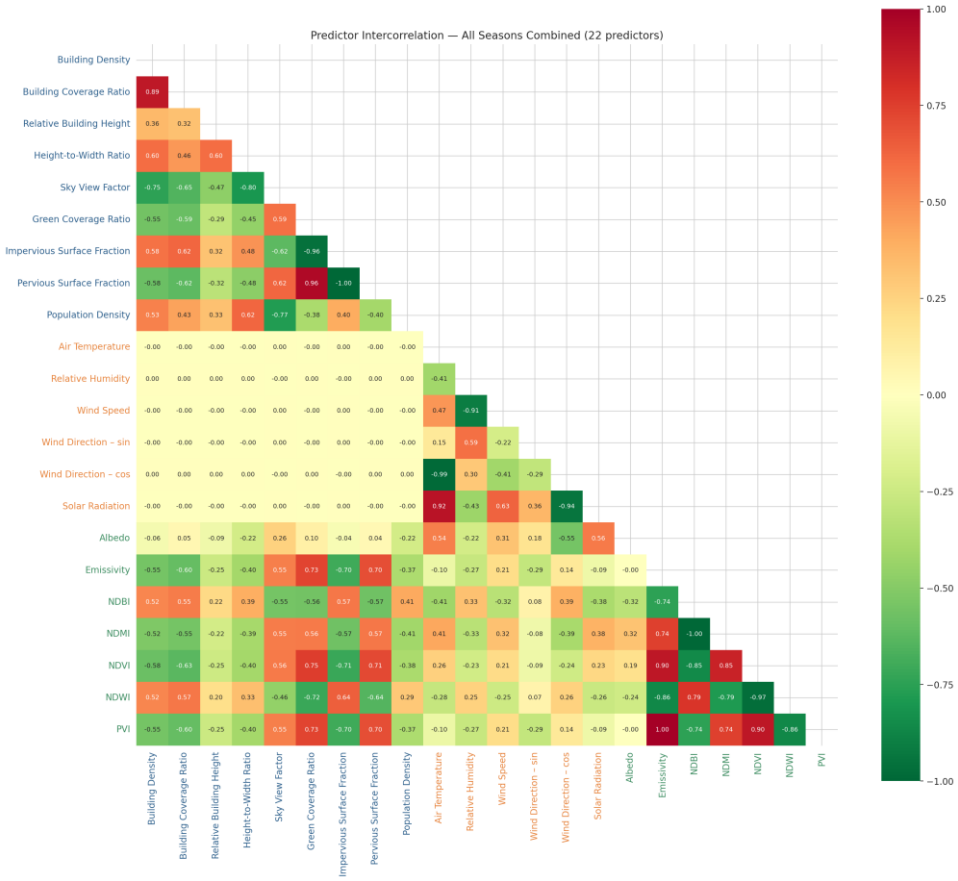


Figure 5.2. Predictor intercorrelation across all twenty-two candidate predictors, all seasons combined.

Seventeen predictor pairs exceed an absolute correlation of 0.85 (Table 5.1), among them several that are algebraically or near-algebraically determined by construction: NDBI and NDMI are exact mirror images (r = −1.00), emissivity and PVI are perfectly collinear (r = 1.00), and impervious and pervious surface fraction are exact complements (r = −1.00). A further cluster of climate variables move together almost perfectly once broadcast across the combined seasonal dataset, air temperature with wind direction cosine (r = −0.99) and with solar radiation (r = 0.92), and relative humidity with wind speed (r = −0.91), reflecting the limited number of distinct seasonal climate states involved, not a genuine spatial relationship. The variance inflation factors confirm that this redundancy extends beyond simple pairwise correlation (Table 5.2): NDBI, wind direction sine

and NDMI each return an infinite VIF once an intercept term is included, and a further twelve predictors exceed the conventional VIF > 10 threshold for severe multicollinearity, most extremely the climate variables and the impervious/pervious surface fraction pair.

Predictor 1	**Predictor 2**	**Pearson r**
NDBI	NDMI	-1.00
Emissivity	PVI	1.00
Impervious Surface Fraction	Pervious Surface Fraction	-1.00
Air Temperature	Wind Direction – cos	-0.99
NDVI	NDWI	-0.97
Green Coverage Ratio	Pervious Surface Fraction	0.96
Green Coverage Ratio	Impervious Surface Fraction	-0.96
Wind Direction – cos	Solar Radiation	-0.94
Air Temperature	Solar Radiation	0.92
Relative Humidity	Wind Speed	-0.91
NDVI	PVI	0.90
Emissivity	NDVI	0.90
Building Density	Building Coverage Ratio	0.89
NDWI	PVI	-0.86
Emissivity	NDWI	-0.86
NDBI	NDVI	-0.85
NDMI	NDVI	0.85

Table 5.1. Predictor pairs exceeding an absolute Pearson correlation of 0.85, all twenty-two candidate predictors, all seasons combined.

Predictor	VIF
NDBI	inf
Wind Direction – sin	inf
NDMI	inf
Wind Speed	3.0×10^{15}
Wind Direction – cos	1.8×10^{15}
Air Temperature	1.0×10^{15}
Relative Humidity	2.3×10^{14}
Solar Radiation	1.0×10^{14}
PVI	5.5×10^{12}
Emissivity	5.5×10^{12}
Pervious Surface Fraction	5.2×10^{8}
Impervious Surface Fraction	5.2×10^{8}
NDVI	108.2
NDWI	36.1
Green Coverage Ratio	20.5

Table 5.2. Predictors exceeding the conventional VIF > 10 threshold before any variable was removed from the twenty-two-predictor candidate set.

On this evidence, thirteen predictors were removed: PVI, emissivity, pervious surface fraction, NDMI, NDWI and green coverage ratio on redundancy grounds, together with NDBI and all six climate predictors. In each collinear pair the retained partner, respectively NDVI in place of PVI and emissivity, impervious surface fraction (ISF) in place of pervious surface fraction, and NDVI in place of NDBI and NDMI, was kept because it carried the more directly interpretable physical meaning for a surface-cover predictor, or because it showed the larger absolute correlation with SUHII in preliminary screening. The climate variables were dropped outright: their extreme VIF values arise mechanically from the

limited number of distinct seasonal states, not from any spatial relationship, and their contribution to every season-specific model was confirmed to be negligible before this final predictor set was adopted, leaving no case for keeping them. Applying the same screening logic within each season separately, in addition to the combined dataset, confirms that the reduction holds consistently: every one of the nine retained predictors stays below the conventional VIF > 10 threshold in every season (Table 5.3), with a maximum across all four seasons of 7.11 (sky view factor, winter).

Predictor	***Mean VIF***	***Max VIF***	***Seasons above 10***
Sky View Factor	6.9936	7.1103	0
Building Density	6.7435	6.7556	0
Building Coverage Ratio	6.4196	6.6053	0
Height-to-Width Ratio	3.5893	3.5902	0
NDVI	2.8076	3.3065	0
Impervious Surface Fraction	2.7285	2.9805	0
Population Density	2.6056	2.6063	0
Relative Building Height	1.6122	1.6164	0
Albedo	1.3213	1.3398	0

Table 5.3. Mean and maximum variance inflation factor for each of the nine predictors in the final set, across the four seasons, and the number of seasons in which each exceeds the conventional VIF > 10 threshold.

The full season-by-season breakdown behind this cross-season summary is given in Table 5.4, confirming that no individual season drives the pattern: every one of the nine predictors stays comfortably below the VIF > 10 threshold in every season, with sky view factor consistently the largest of the nine, ranging from 6.92 in summer to 7.14 in winter.

Season	*Predictor*	*VIF*
Summer	Sky View Factor	6.9159
	Building Density	6.7556
	Building Coverage Ratio	6.4313
	Height-to-Width Ratio	3.5896
	NDVI	3.3065
	Impervious Surface Fraction	2.9805
	Population Density	2.6052
	Relative Building Height	1.6083
	Albedo	1.3133
Winter	Sky View Factor	7.1103
	Building Density	6.7231
	Building Coverage Ratio	6.2252
	Height-to-Width Ratio	3.5885
	Impervious Surface Fraction	2.6069
	Population Density	2.6057
	NDVI	2.3835
	Relative Building Height	1.6164
	Albedo	1.3398
Spring	Sky View Factor	6.9084
	Building Density	6.7397
	Building Coverage Ratio	6.4166
	Height-to-Width Ratio	3.5888
	Impervious Surface Fraction	2.6208
	Population Density	2.6053
	NDVI	2.5201
	Relative Building Height	1.6163
	Albedo	1.3150
Autumn	Sky View Factor	7.0399
	Building Density	6.7554
	Building Coverage Ratio	6.6053
	Height-to-Width Ratio	3.5902
	NDVI	3.0204
	Impervious Surface Fraction	2.7057
	Population Density	2.6063
	Relative Building Height	1.6076
	Albedo	1.3171

Table 5.4. Post-reduction variance inflation factor for each of the nine predictors in the final set, by season.

5.5 Random Forest Modelling

The final nine predictors established in Sections 5.3 and 5.4 are used to develop the four seasonal Random Forest models reported in this chapter, each fitted independently on that season's grid cells alone. A fifth configuration pools all four seasons into a single 53,785-row sample and retains the fifteen-predictor set, climate included, since combining the four seasons is the only setting in which the climate predictors carry genuine spatial variance for the model to learn from. Since SUHI and SUHII remain distinct response variables throughout, each of these five configurations produces two fitted models in practice, one per target. Tuning was carried out configuration by configuration, with RandomizedSearchCV fitting RandomForestRegressor separately for SUHII and SUHI, scored under five-fold cross-validation using R^2. Every dataset went through the same 80:20 training/test division, governed by a fixed random seed. The best configuration from each search was refit on the full training set to produce the final model, with the selected hyperparameters reported in Table 5.5.

5.6 Model Performance and Validation

Model performance was assessed on the held-out 20% test set using the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE), and cross-checked against five-fold cross-validated R^2 computed on the full dataset, giving a sense of how a single test-set score compares against the wider range of performance cross-validation reveals. Table 5.6 lays out all three measures across every season, the All Seasons Combined configuration, and both targets. Random Forest tracks the observed SUHII and SUHI values closely throughout: test-set R^2 reaches 0.90 in summer for the seasonal models and stays above 0.80 for both targets in the All Seasons

Combined model, with winter as the low point at 0.70. The cross-validated R^2 sits lower than the test-set score across every row, with the widest spread in winter (0.462 ± 0.282). This points to some cross-validation folds spanning grid cells with fairly different urban contexts.

Configuration	n_estimators	max_depth	min_samples_leaf	max_features
Summer	300	16	3	log2
Winter	300	12	5	0.5
Spring	200	16	3	log2
Autumn	300	16	3	log2
All Seasons Combined	300	12	5	0.5

Table 5.5. Random Forest hyperparameters selected by randomised search for each configuration, identical for SUHI and SUHII. Configurations use the nine-predictor set, except All Seasons Combined, which adds the six climate predictors, since pooling the four seasons is the only setting where they carry genuine spatial variance.

Configuration	Target	Test R^2	RMSE	MAE	CV R^2 (mean ± sd)
Summer	SUHII	0.902	0.344	0.252	0.713 ± 0.185
Summer	SUHI	0.902	1.681	1.233	0.713 ± 0.185
Winter	SUHII	0.701	0.581	0.402	0.462 ± 0.282
Winter	SUHI	0.701	0.850	0.590	0.461 ± 0.282
Spring	SUHII	0.807	0.508	0.361	0.487 ± 0.331
Spring	SUHI	0.807	1.059	0.754	0.487 ± 0.331
Autumn	SUHII	0.809	0.475	0.342	0.554 ± 0.272
Autumn	SUHI	0.809	1.243	0.895	0.554 ± 0.272
All Seasons Combined	SUHII	0.799	0.508	0.347	0.618 ± 0.087
All Seasons Combined	SUHI	0.893	1.309	0.891	0.543 ± 0.226

Table 5.6. Random Forest test-set and five-fold cross-validated R^2 for every season, the All Seasons Combined configuration, and both targets. Rows use the nine-predictor model, except the combined configuration, which uses the fifteen-predictor model.

Summer gives the best model fit among the seasonal models for both targets, with a test R^2 of 0.902. Winter gives the weakest fit, at 0.701, with spring and autumn falling in between at 0.807 and 0.809 respectively. This order fits the seasonal spread of spectral predictors from Chapters 3 and 4: when a season has a narrow range of spectral signal, the Random Forest has less variation to work with, and that is a limitation of the modelling itself. The All Seasons Combined model sits within this same range for SUHII (0.799) and slightly above the best seasonal score for SUHI (0.893), indicating that pooling the four seasons does not come at a material cost to predictive accuracy despite the wider variety of conditions the pooled model has to cover.

Residual diagnostics, actual versus predicted scatter, residuals against predicted value, and the distribution of residuals against a fitted normal curve, were produced for every season, the All Seasons Combined configuration, and both targets. Figures 5.3 to 5.6 show the SUHII diagnostics for each season in turn; in every case the mean residual sits close to zero and no strong systematic curvature appears against the predicted value, with a modest widening of the residual spread toward the extremes of the predicted range, most visible in winter and spring, the two seasons with the weakest overall fit. Figures 5.7 and 5.8 show the equivalent diagnostics for the All Seasons Combined model, for SUHII and SUHI respectively, and show the same pattern of well-centred residuals without systematic curvature.

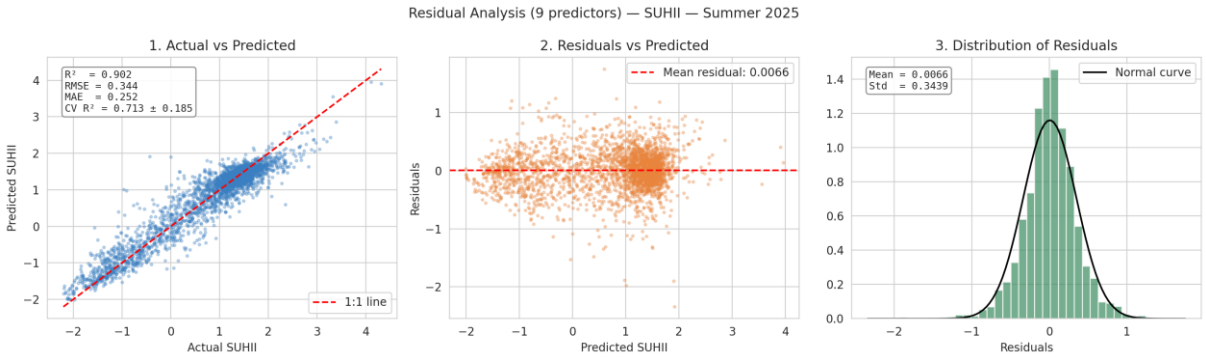


Figure 5.3. Residual diagnostics, summer SUHII: actual versus predicted, residuals versus predicted, and the distribution of residuals.

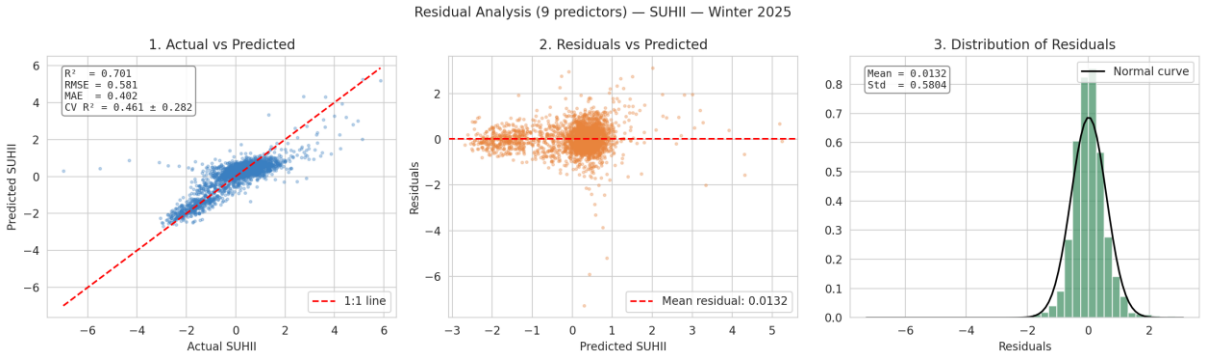


Figure 5.4. Residual diagnostics, winter SUHII: actual versus predicted, residuals versus predicted, and the distribution of residuals.

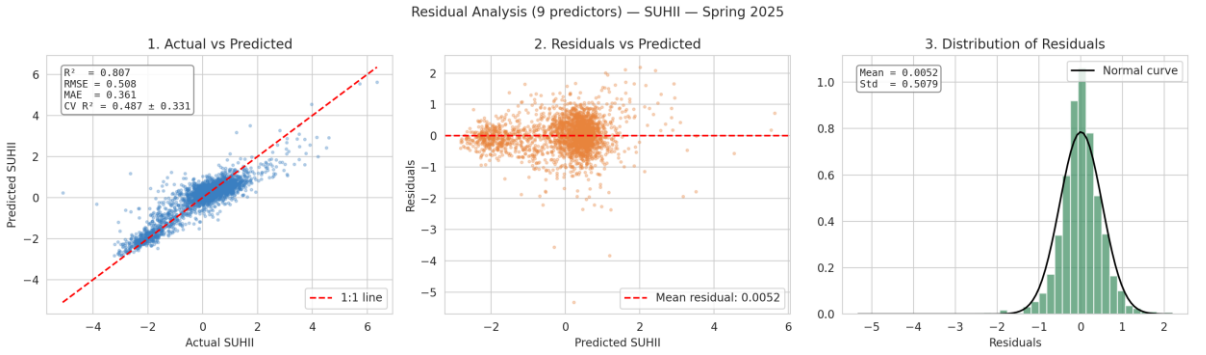


Figure 5.5. Residual diagnostics, spring SUHII: actual versus predicted, residuals versus predicted, and the distribution of residuals.

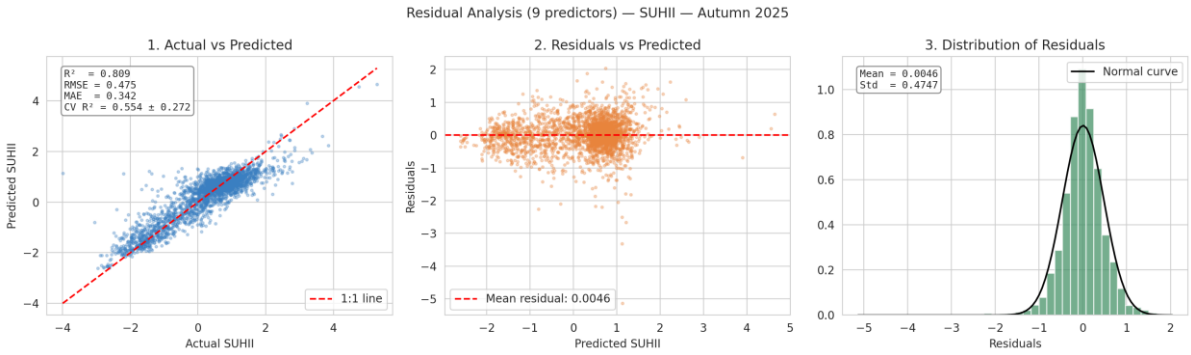


Figure 5.6. Residual diagnostics, autumn SUHII: actual versus predicted, residuals versus predicted, and the distribution of residuals.

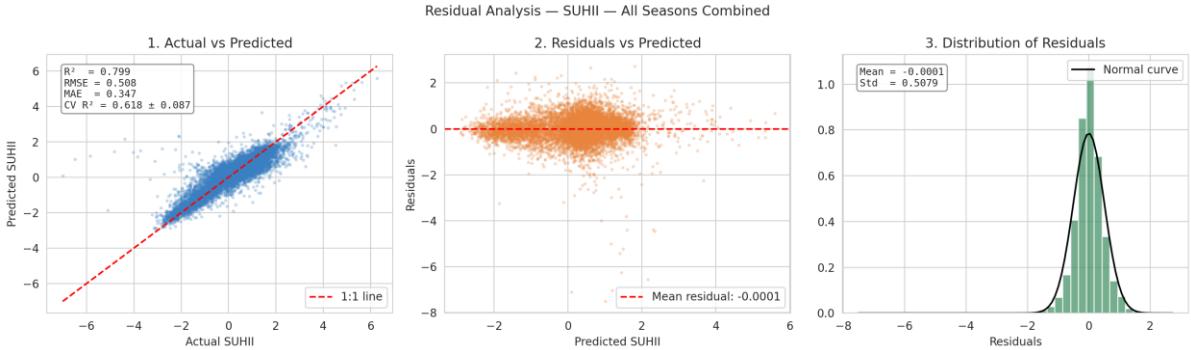


Figure 5.7. Residual diagnostics, All Seasons Combined SUHII.

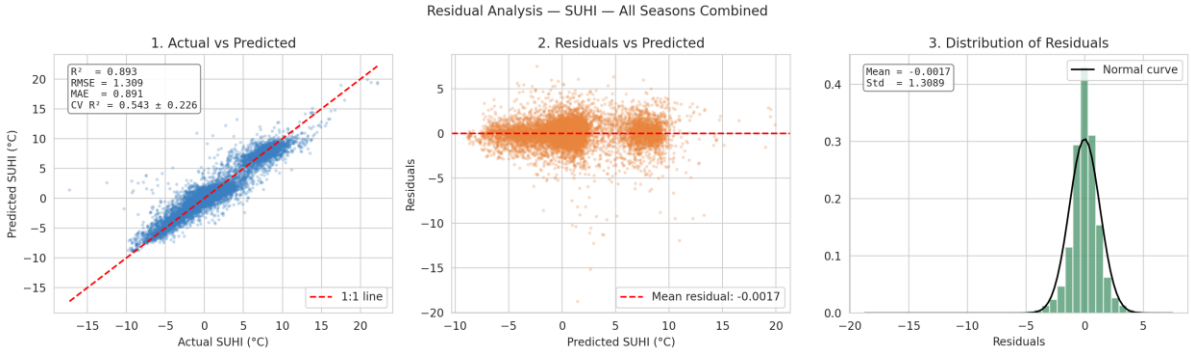


Figure 5.8. Residual diagnostics, All Seasons Combined SUHI.

5.7 Feature Importance and SHAP Interpretation

Two complementary measures are used to interpret the fitted models. Gini feature importance, aggregated from the fitted trees, gives a globally averaged measure of each predictor's contribution to reducing prediction error across the whole dataset. SHAP values decompose each individual grid cell's prediction into additive per-predictor contributions, so that both the size and the direction

of a predictor's pull on a given prediction can be examined, and are visualised here as beeswarm plots, in which each point is one grid cell coloured by that cell's value for the predictor concerned. Both measures describe the contribution of a predictor to the model's output; neither measure, nor the Random Forest itself, establishes that a predictor causes the observed variation in SUHI or SUHII, and the discussion in this chapter is accordingly phrased in terms of modelled association, contribution and importance within the fitted model, without any claim to a causal effect.

5.7.1 Summer 2025

NDVI dominates summer SUHII importance, accounting for close to half of the total (0.448), with impervious surface fraction a distant second (0.194) and building density third (0.096). A one-standard-deviation increase in NDVI is associated with the largest model-implied change in SUHII (−0.620) and SUHI (−3.033 °C) of any predictor, consistent with its strongly negative Pearson correlation. Four predictors are flagged for a sign flip between their marginal sensitivity and their raw correlation: albedo, height-to-width ratio, relative building height and sky view factor, indicating that their apparent univariate relationship with SUHII reverses once the other eight predictors are held fixed.

Predictor	*Importance — SUHII*	*Importance — SUHI*
NDVI	0.4478	0.4479
Impervious Surface Fraction	0.1937	0.1937
Building Density	0.0956	0.0955
Building Coverage Ratio	0.0828	0.083
Albedo	0.0731	0.0731
Sky View Factor	0.0438	0.0437
Population Density	0.0382	0.0381
Height-to-Width Ratio	0.0136	0.0137
Relative Building Height	0.0114	0.0114

Table 5.7. Feature importance for summer SUHII, with the matching SUHI figures; all nine predictors.

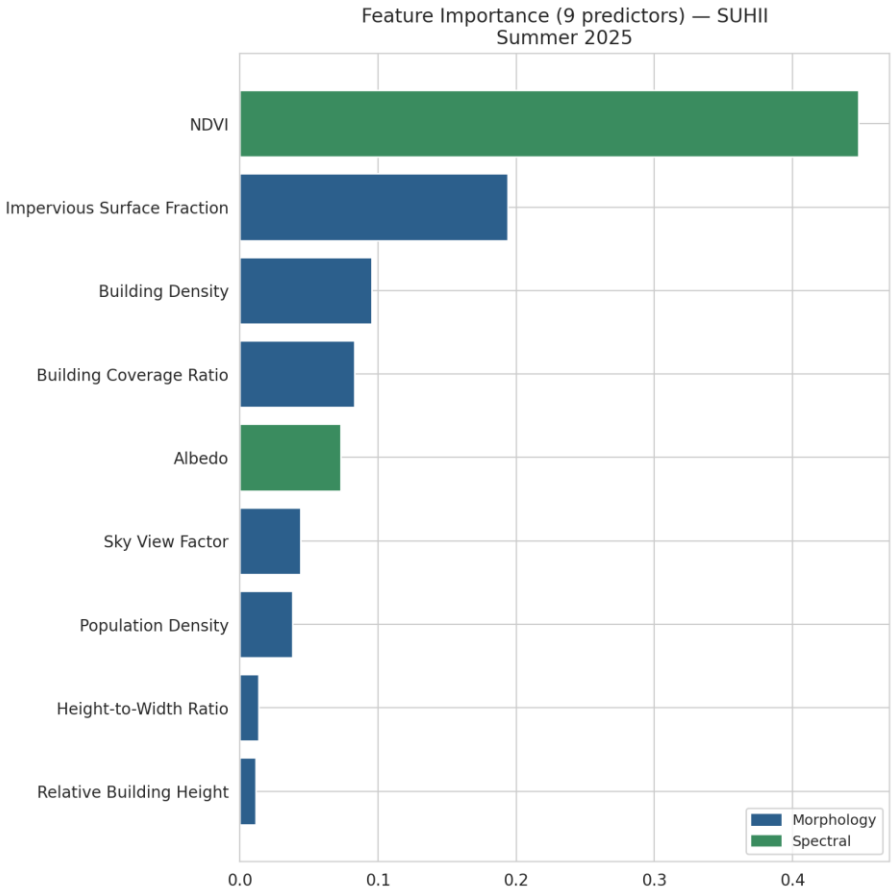


Figure 5.9. Feature importance for summer SUHII.

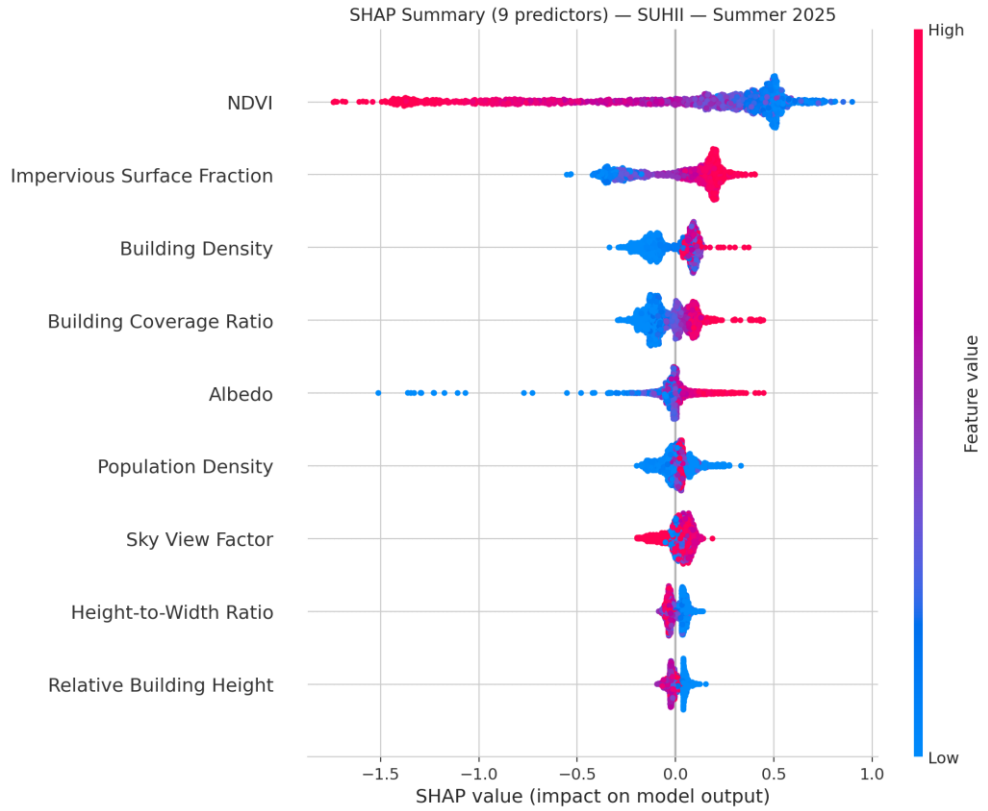


Figure 5.10. SHAP beeswarm for summer SUHII, showing how each predictor pushes individual grid-cell predictions up or down.

Predictor	***Δ SUHII***	***Δ SUHI***	***Pearson r (SUHII)***	***Sign flip***
NDVI	-0.6202	-3.0328	-0.8629	No
Building Coverage Ratio	0.0939	0.46	0.5809	No
Building Density	0.0858	0.4226	0.5222	No
Impervious Surface Fraction	0.0742	0.362	0.7404	No
Albedo	0.0649	0.3168	-0.0326	Yes
Population Density	0.0484	0.2364	0.3243	No
Height-to-Width Ratio	-0.0319	-0.1552	0.3548	Yes
Relative Building Height	-0.0038	-0.0179	0.2461	Yes
Sky View Factor	0.0018	0.0104	-0.4906	Yes

Table 5.8. One-at-a-time (+1 SD) sensitivity for summer, with the sign-flip flag against each predictor's raw Pearson correlation.

5.7.2 Winter 2025

NDVI (0.316) and albedo (0.270) are jointly the leading predictors of winter SUHII, together accounting for close to three-fifths of summed importance; impervious surface fraction (0.111) and building coverage ratio (0.094) follow behind. Four predictors are flagged for a sign flip: sky view factor, population density, relative building height and height-to-width ratio, matching the largest count of any season and consistent with winter's weaker overall thermal signal leaving more predictors vulnerable to confounding once the full retained set is held fixed.

Predictor	***Importance — SUHII***	***Importance — SUHI***
NDVI	0.3157	0.3159
Albedo	0.2696	0.2695
Impervious Surface Fraction	0.1109	0.111
Building Coverage Ratio	0.094	0.0942
Building Density	0.0764	0.0762
Population Density	0.0451	0.0451
Sky View Factor	0.0365	0.0365
Height-to-Width Ratio	0.0273	0.0272
Relative Building Height	0.0245	0.0245

Table 5.9. Feature importance for winter SUHII, with the matching SUHI figures; all nine predictors.

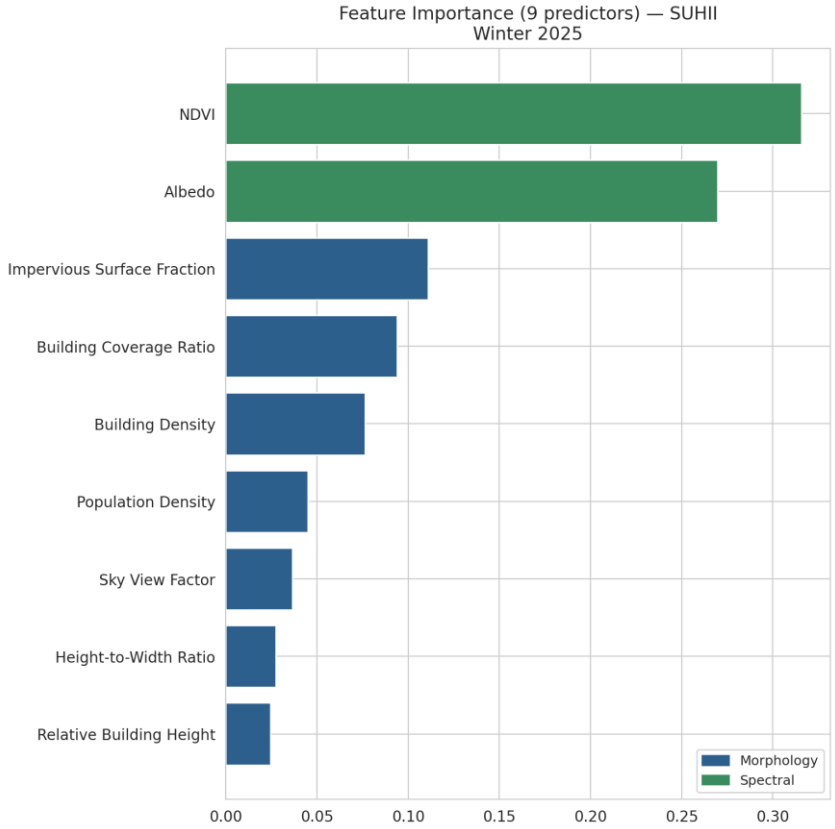


Figure 5.11. Feature importance for winter SUHII.

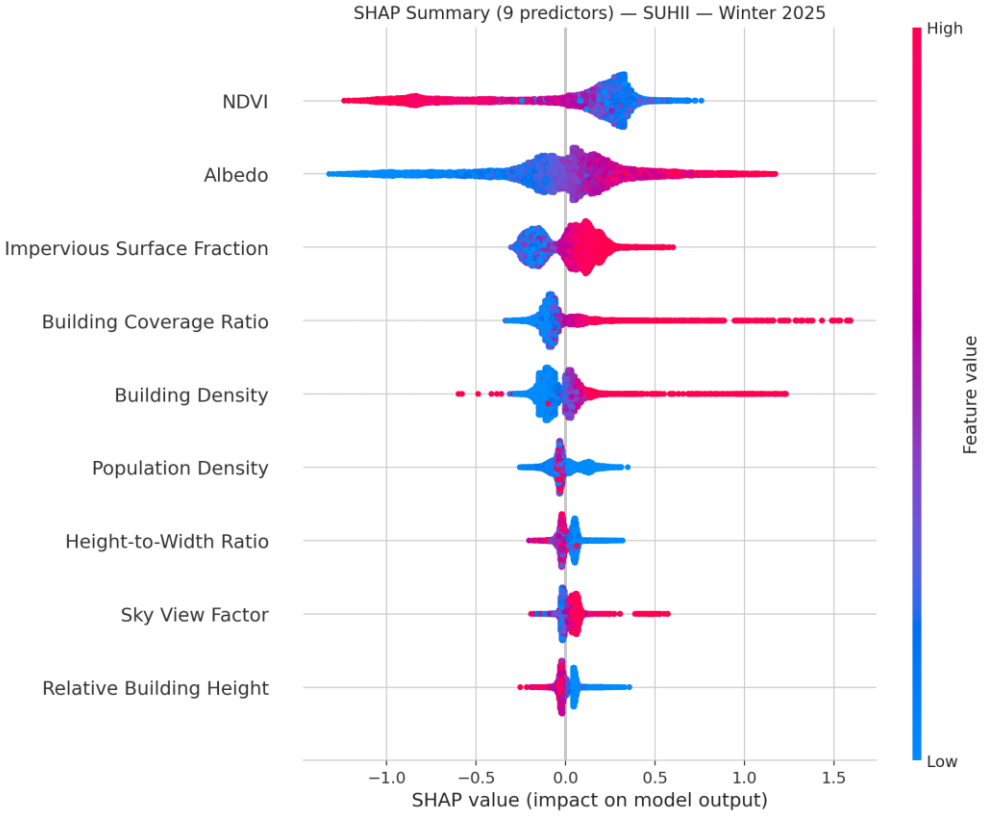


Figure 5.12. SHAP beeswarm for winter SUHII, showing how each predictor pushes individual grid-cell predictions up or down.

Predictor	*Δ SUHII*	*Δ SUHI*	*Pearson r (SUHII)*	*Sign flip*
NDVI	-0.3161	-0.4634	-0.5781	No
Albedo	0.2878	0.4219	0.3876	No
Building Density	0.1616	0.2355	0.4221	No
Building Coverage Ratio	0.0995	0.1458	0.4495	No
Impervious Surface Fraction	0.074	0.1085	0.476	No
Sky View Factor	0.0307	0.0447	-0.2428	Yes
Population Density	-0.0264	-0.039	0.1499	Yes
Relative Building Height	-0.0223	-0.0327	0.0696	Yes
Height-to-Width Ratio	-0.0212	-0.031	0.1435	Yes

Table 5.10. One-at-a-time (+1 SD) sensitivity for winter, with the sign-flip flag against each predictor's raw Pearson correlation.

5.7.3 Spring 2025

NDVI (0.270) again leads spring SUHII importance, followed by albedo (0.218) and impervious surface fraction (0.145), with building coverage ratio (0.106) and building density (0.097) as the next-largest contributors. Two predictors are flagged for a sign flip: height-to-width ratio and sky view factor, tied with autumn for the fewest of the four seasons.

Predictor	*Importance — SUHII*	*Importance — SUHI*
NDVI	0.2699	0.2699
Albedo	0.218	0.2184
Impervious Surface Fraction	0.1445	0.1442
Building Coverage Ratio	0.1064	0.1064
Building Density	0.0968	0.0969
Population Density	0.065	0.0648
Sky View Factor	0.0478	0.0477
Height-to-Width Ratio	0.0263	0.0262
Relative Building Height	0.0254	0.0255

Table 5.11. Feature importance for spring SUHII, with the matching SUHI figures; all nine predictors.

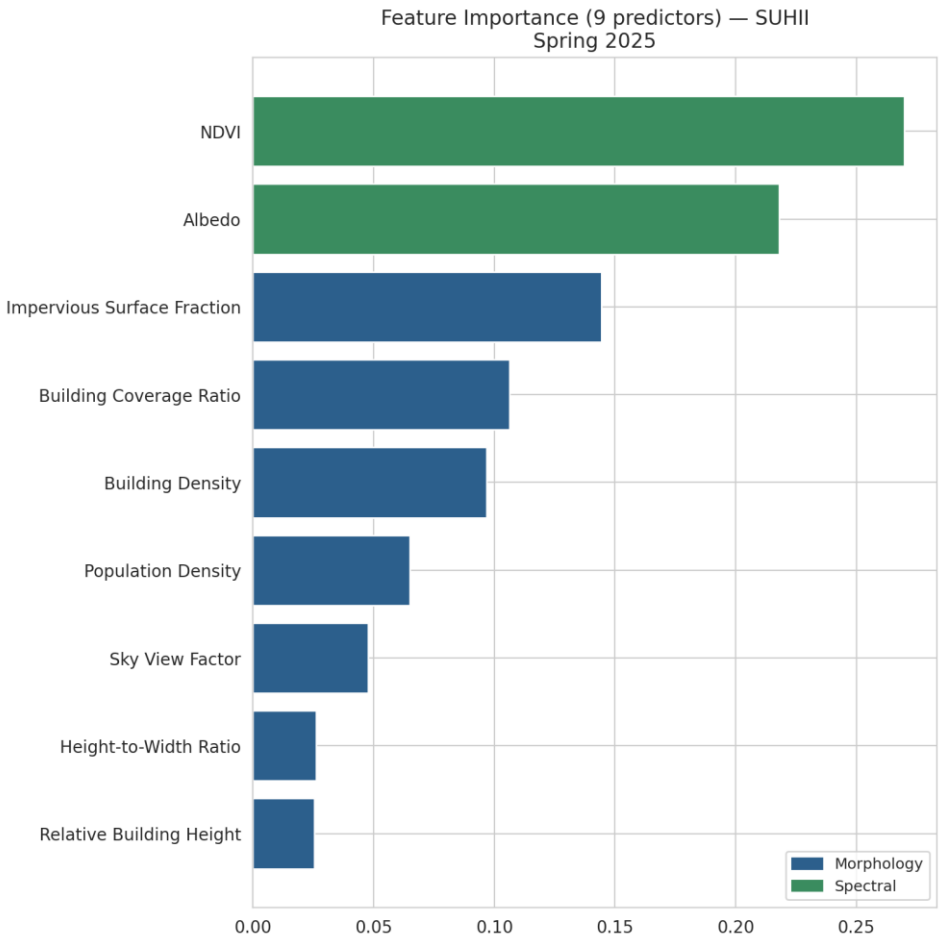


Figure 5.13. Feature importance for spring SUHII.

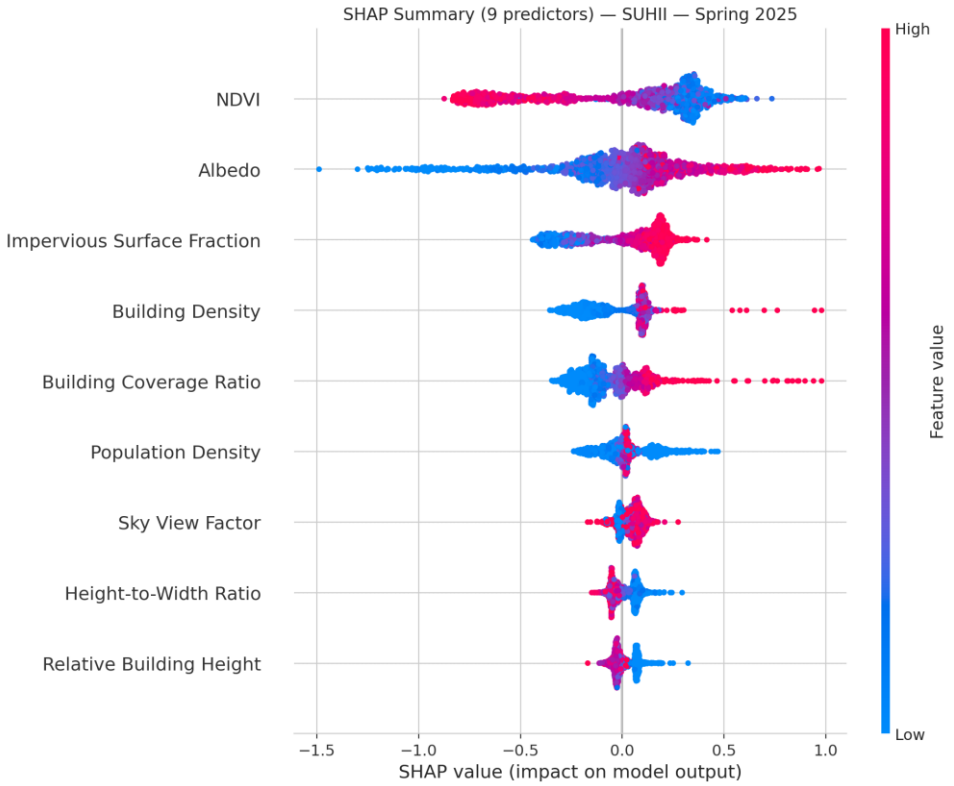


Figure 5.14. SHAP beeswarm for spring SUHII, showing how each predictor pushes individual grid-cell predictions up or down.

Predictor	***Δ SUHII***	***Δ SUHI***	***Pearson r (SUHII)***	***Sign flip***
NDVI	-0.3085	-0.6441	-0.5979	No
Albedo	0.2572	0.5372	0.352	No
Building Density	0.2135	0.448	0.4878	No
Building Coverage Ratio	0.1986	0.4144	0.5398	No
Impervious Surface Fraction	0.1067	0.223	0.6237	No
Population Density	0.059	0.1243	0.2261	No
Height-to-Width Ratio	-0.0404	-0.0838	0.2326	Yes
Sky View Factor	0.0154	0.0303	-0.353	Yes
Relative Building Height	0.0002	0.0003	0.1583	No

Table 5.12. One-at-a-time (+1 SD) sensitivity for spring, with the sign-flip flag against each predictor's raw Pearson correlation.

5.7.4 Autumn 2025

NDVI holds the highest importance weight in autumn at 0.372, with impervious surface fraction (0.184) and albedo (0.126) occupying the next positions along the importance continuum. Sign reversals are recorded for two predictors, height-to-width ratio and sky view factor, tied with spring for the lowest count across the four seasons, reflecting that autumn's predictor–target relationships sit close to internal consistency once the full retained set is held fixed.

Predictor	***Importance — SUHII***	***Importance — SUHI***
NDVI	0.3717	0.3718
Impervious Surface Fraction	0.1838	0.1838
Albedo	0.126	0.126
Building Density	0.0897	0.0897
Building Coverage Ratio	0.0809	0.0808
Population Density	0.0539	0.0538
Sky View Factor	0.0493	0.0494
Height-to-Width Ratio	0.0241	0.0242
Relative Building Height	0.0206	0.0206

Table 5.13. Feature importance for autumn SUHII, with the matching SUHI figures; all nine predictors.

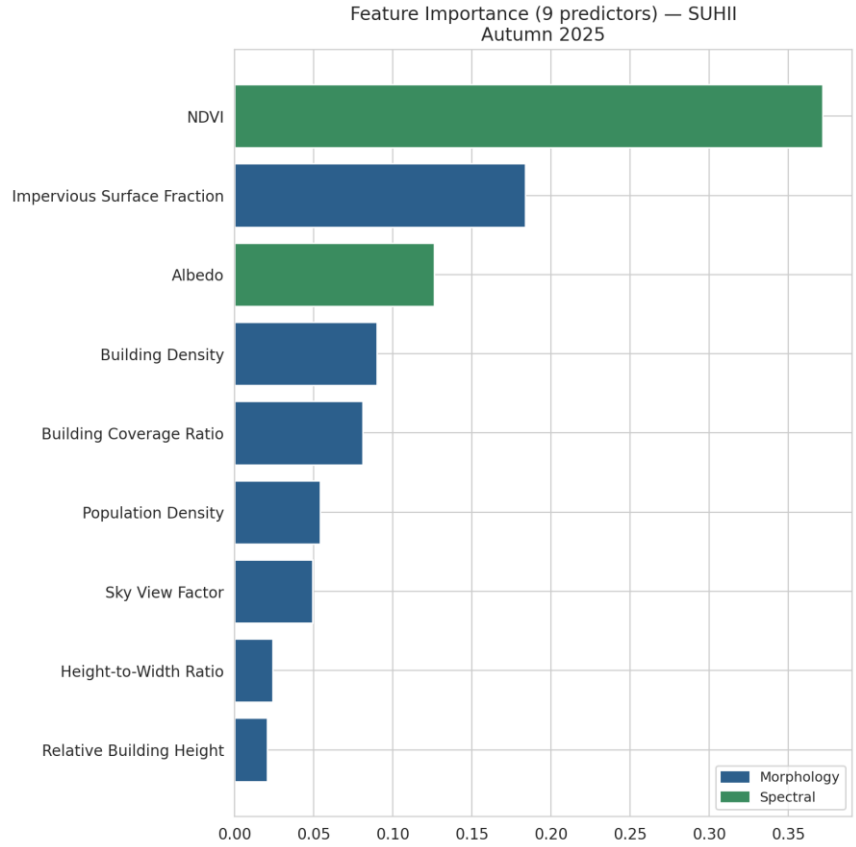


Figure 5.15. Feature importance for autumn SUHII.

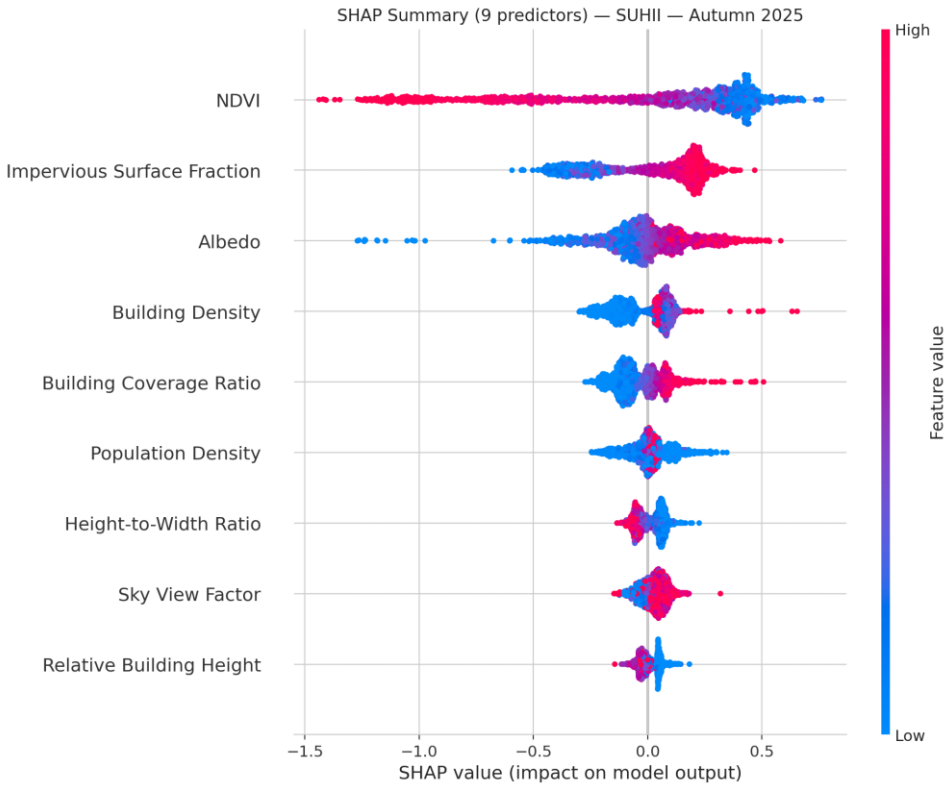


Figure 5.16. SHAP beeswarm for autumn SUHII, showing how each predictor pushes individual grid-cell predictions up or down.

Predictor	*Δ SUHII*	*Δ SUHI*	*Pearson r (SUHII)*	*Sign flip*
NDVI	-0.4706	-1.232	-0.712	No
Building Density	0.1454	0.3822	0.4611	No
Albedo	0.1425	0.373	0.1899	No
Building Coverage Ratio	0.1282	0.337	0.5269	No
Impervious Surface Fraction	0.1183	0.3103	0.6624	No
Height-to-Width Ratio	-0.0518	-0.1364	0.254	Yes
Sky View Factor	0.0484	0.1274	-0.3728	Yes
Population Density	0.0361	0.0947	0.238	No
Relative Building Height	0.0056	0.0149	0.1867	No

Table 5.14. One-at-a-time (+1 SD) sensitivity for autumn, with the sign-flip flag against each predictor's raw Pearson correlation.

5.7.5 All Seasons Combined

NDVI leads All Seasons Combined importance for both targets (0.258 for SUHII, 0.368 for SUHI), with impervious surface fraction (0.249, 0.072) and albedo (0.153, 0.075) following; unlike the season-specific models, the climate predictors now carry real, if modest, importance, since combining the four seasons gives them genuine row-to-row variance for the first time. Wind direction cosine, solar radiation and air temperature each contribute markedly more to SUHI importance than to SUHII importance (Table 5.15), a pattern confirmed by the mean absolute SHAP values (Table 5.16), where the same three climate predictors carry mean |SHAP| values of 0.520, 0.627 and 0.467 respectively for SUHI, against 0.061, 0.054 and 0.057 for SUHII.

Predictor	Importance — SUHII	Importance — SUHI
NDVI	0.258	0.368
Impervious Surface Fraction	0.249	0.072
Albedo	0.153	0.075
Building Density	0.070	0.034
Building Coverage Ratio	0.057	0.030
Population Density	0.044	0.022
Wind Direction – cos	0.029	0.108
Sky View Factor	0.029	0.015
Air Temperature	0.027	0.101
Solar Radiation	0.026	0.129
Height-to-Width Ratio	0.016	0.009
Relative Building Height	0.015	0.007
Relative Humidity	0.010	0.010
Wind Direction – sin	0.010	0.013
Wind Speed	0.005	0.008

Table 5.15. Feature importance for the All Seasons Combined model, both targets.

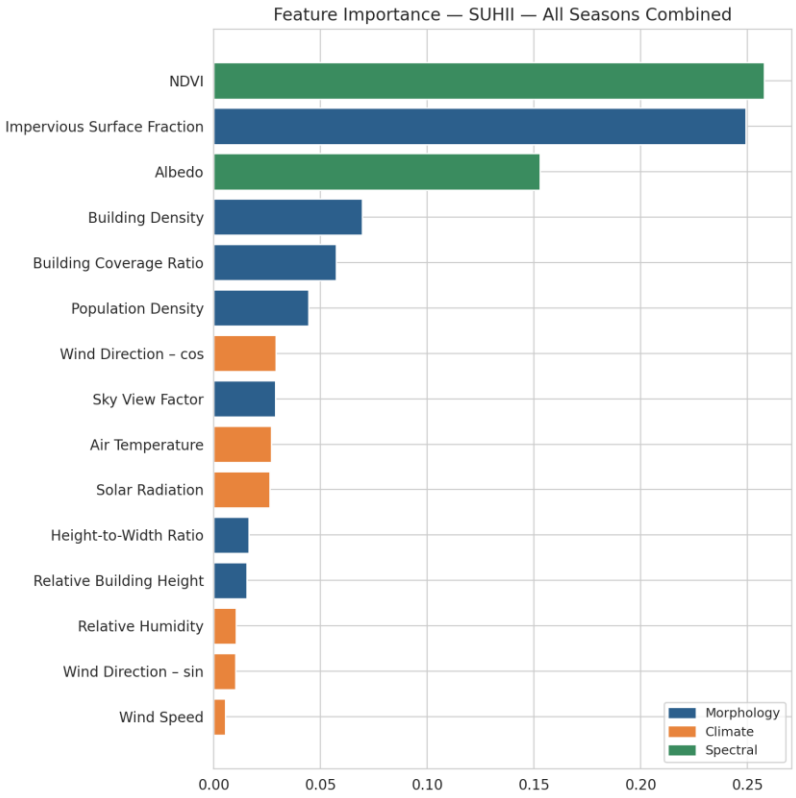


Figure 5.17. Feature importance for the All Seasons Combined SUHII model.

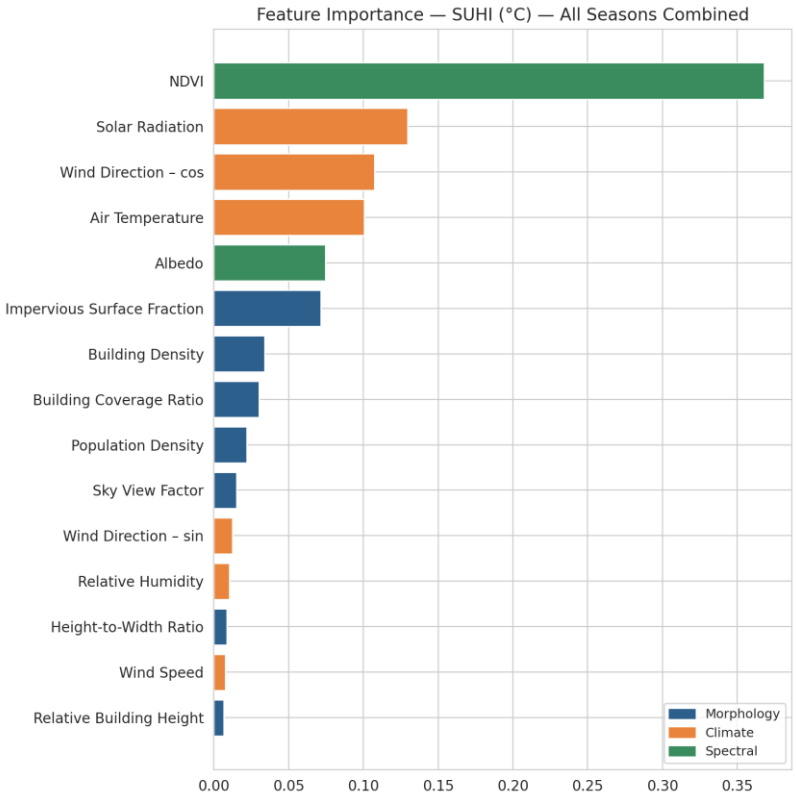


Figure 5.18. Feature importance for the All Seasons Combined SUHI model.

Predictor	Mean |SHAP| SUHII	Mean |SHAP| SUHI
NDVI	0.370	1.524
Impervious Surface Fraction	0.272	0.381
Albedo	0.239	0.561
Building Density	0.099	0.170
Building Coverage Ratio	0.082	0.191
Population Density	0.062	0.146
Wind Direction – cos	0.061	0.520
Air Temperature	0.057	0.467
Solar Radiation	0.054	0.627
Height-to-Width Ratio	0.042	0.114
Sky View Factor	0.038	0.102
Wind Direction – sin	0.030	0.100
Relative Building Height	0.028	0.068
Relative Humidity	0.022	0.105
Wind Speed	0.013	0.099

Table 5.16. Mean absolute SHAP value for each predictor in the All Seasons Combined model, both targets.

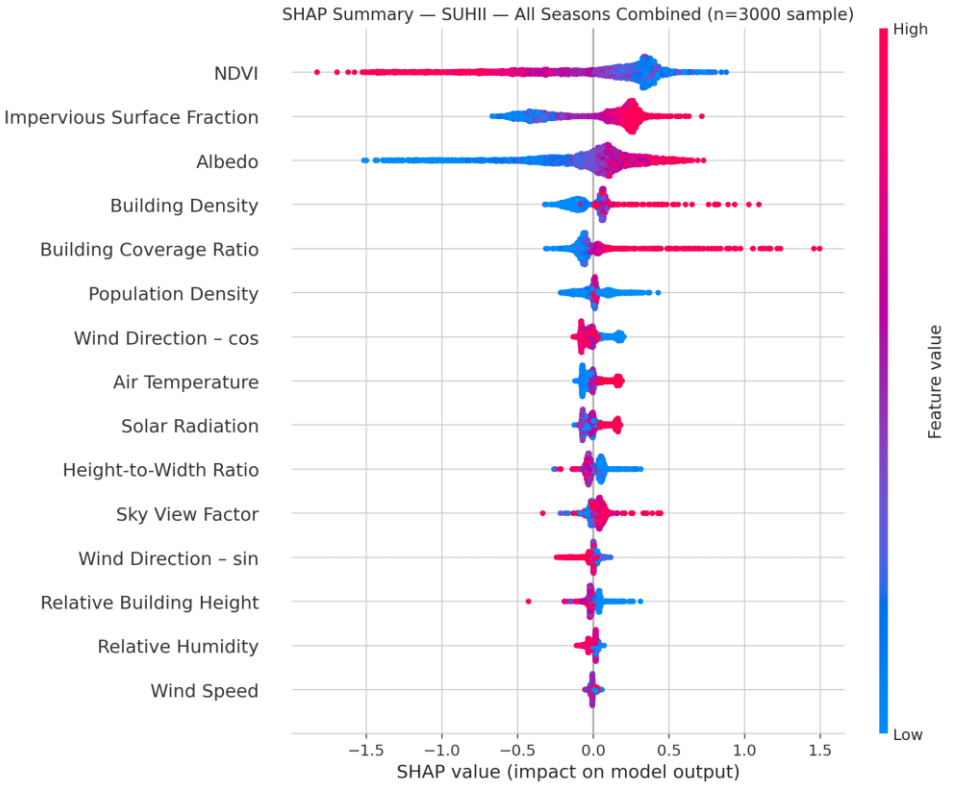


Figure 5.19. SHAP beeswarm for the All Seasons Combined SUHII model.

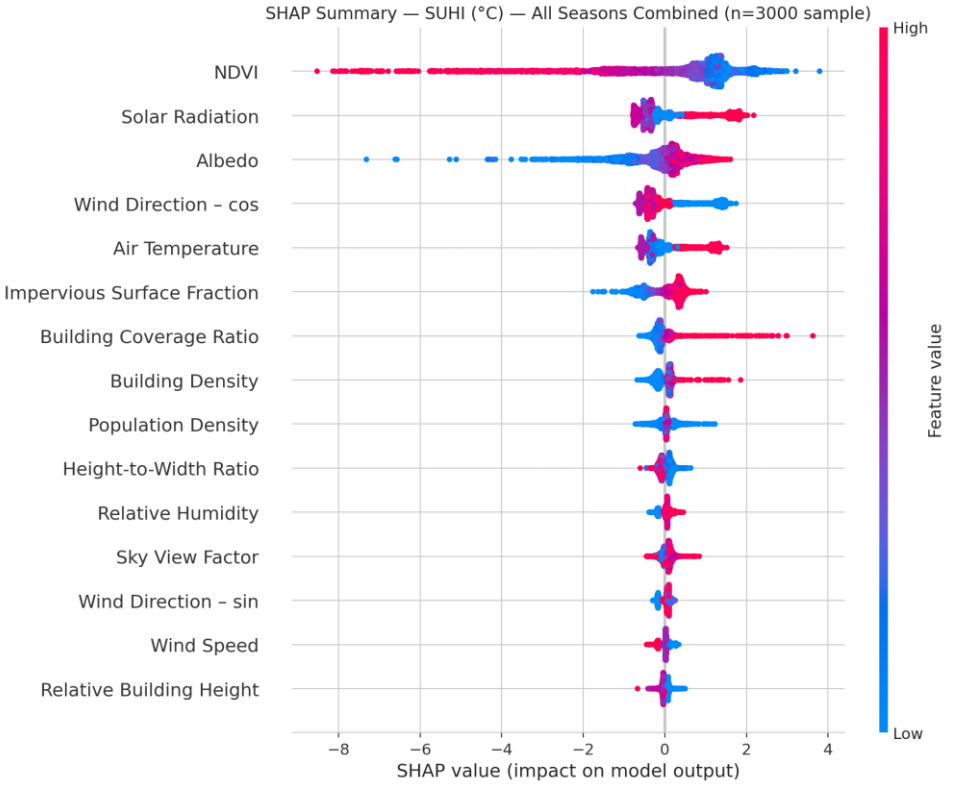


Figure 5.20. SHAP beeswarm for the All Seasons Combined SUHI model.

Table 5.17 lays the climate–SUHI relationship. For SUHII, morphology leads at 48%, spectral condition follows close behind at 41%, and climate trails at just 11%. For SUHI, the picture changes: spectral condition still leads at 44%, but climate jumps to 37%, over three times its SUHII share while morphology drops to 19%. This fits what is expected from SUHI's definition: since it's an absolute temperature departure, it responds more directly to the day's ambient thermal and radiative conditions than SUHII's rural-normalised measure does.

Group	Share — SUHII	Share — SUHI
Morphology	48%	19%
Climate	11%	37%
Spectral	41%	44%

Table 5.17. Predictor-group contribution to summed importance in the All Seasons Combined model, both targets.

Five predictors are flagged for a sign flip between their marginal sensitivity and their raw Pearson correlation in the All Seasons Combined model (Table 5.18): height-to-width ratio, sky view factor, wind direction sine, relative building height and relative humidity. All five also appear among the sign-flip predictors in at least one of the four seasonal models, reinforcing the conclusion that these particular morphological, directional and climatic variables carry a confounded relationship with surface heat once the remaining predictors are held fixed.

Predictor	Δ SUHII per +1 SD	Pearson r (SUHII)
Height-to-Width Ratio	-0.031	0.24
Sky View Factor	0.019	-0.35
Wind Direction – sin	-0.016	0.07
Relative Building Height	-0.009	0.16
Relative Humidity	0.006	-0.04

Table 5.18. The five predictors whose marginal sensitivity flips sign against their raw Pearson correlation in the All Seasons Combined model.

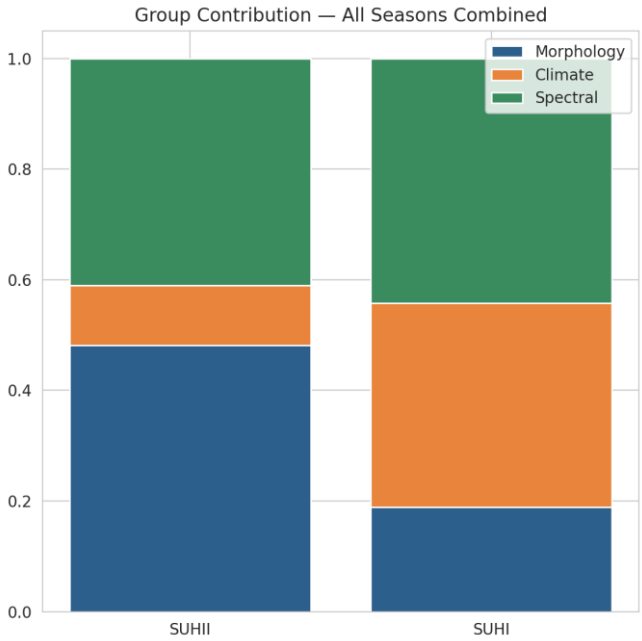


Figure 5.21. Predictor-group contribution to SUHII and SUHI, All Seasons Combined model.

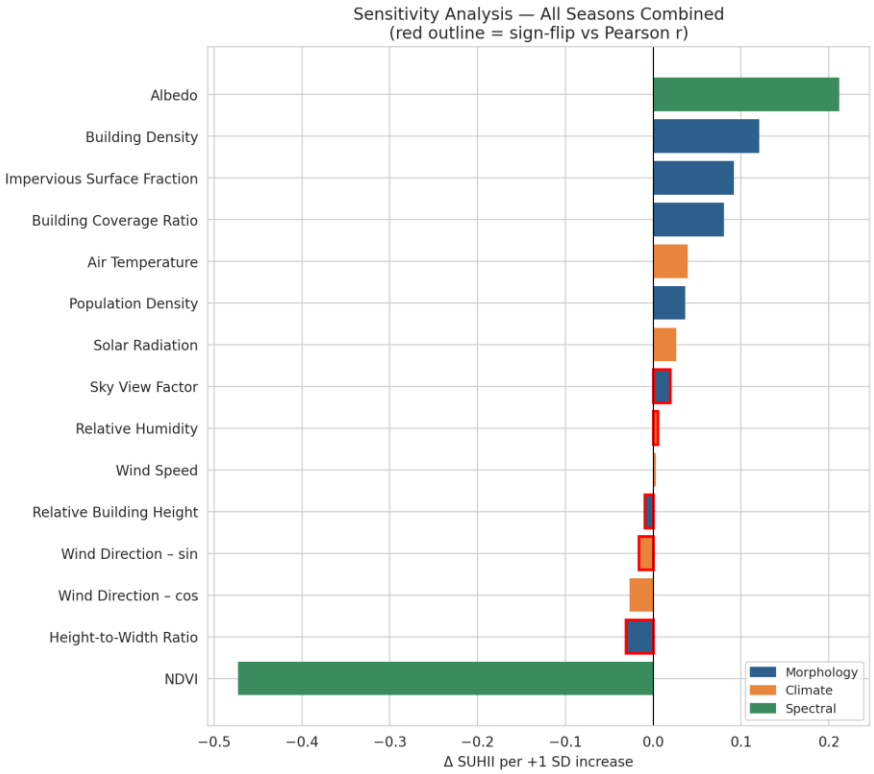


Figure 5.22. Model-implied sensitivity of All Seasons Combined SUHII to a one-standard-deviation change in each predictor.

5.8 Seasonal and Cross-Season Interpretation

Across all four seasons, NDVI leads Gini feature importance for SUHII without exception, ranging from a maximum of 0.448 in summer down to 0.270 in spring, while albedo consistently ranks second in the three cooler seasons, peaking in winter (0.270) and falling behind impervious surface fraction into third place in summer (0.073) and autumn (0.126). Impervious surface fraction and building density or building coverage ratio emerge as the leading remaining contributors across all four seasons, with the ordering among these three shifting slightly from season to season. NDVI alone carries most of the spectral signal in every season, since the algebraically related spectral predictors removed in Section 5.4 are no longer available to redistribute importance between seasons, and since climate is excluded from this predictor set entirely, morphology and the two retained

spectral predictors between them account for the whole of summed importance in every season.

Pooling the four seasons into the All Seasons Combined model changes this picture: with climate finally free to vary from row to row, morphology and spectral condition arrive at similar shares of summed SUHII importance (48% and 41%), with climate contributing a modest but non-zero 11%, and climate's share for SUHI rising further still, to 37%. This indicates that each seasonal model's zero climate importance in Section 5.7 follows from how those models are built, holding one season's climate fixed; climate itself retains a physical role in shaping surface heat, visible only once the four seasons are pooled.

Within each season, SUHI and SUHII have the same predictor ranking. Section 5.7 makes this clear, with SUHI and SUHII figures tracking each other closely throughout. For example, NDVI moves SUHI several times more than SUHII in absolute terms because SUHI is measured in wider units (°C). Height-to-width ratio and sky view factor show a sign flip between their correlation and their model-based sensitivity in every season (2 to 4 predictors flip overall, depending on the season). This means these two variables can look misleading if read alone, without the full model.

5.9 Predicted SUHII and SUHI

To make the seasonal SUHII predictions legible at the scale the Comune di Torino uses for planning and statistical reporting, the 100 x100 m grid output was aggregated to the city's Zone Statistiche polygons. Published by Comune di Torino through the aperTO open-data portal, the final statistical zones were used. Figure 5.23 to Figure 5.26 present the resulting zone-level SUHII predictions for summer, winter, spring and autumn. The hill side district on the eastern side of the

Po River remains to be the coolest part of the city in every season. Whereas the historic core and the northern industrial fringe register the warmest values throughout the year. One zone in the city's southwest (Nuovo Fiat- statistical zone) corner reaches the extreme hotspot class in all four seasons map. This tells that it is a persistent outlier independent of season. The extent of the hottest classes grows from winter to summer. In winter, most of the built-up area sits within moderate to hot range and extreme-hotspot is limited to two zones. Whereas in summer, nearly entire built-up area- from the historic core through the western and northern districts, falls into the very-hot to extreme- hotspot class. Spring and Autumn occupy an intermediate position between the two extremes. Autumn shows visibly wider spread of hot-zones than spring (figure 5.23), particularly along the northern boundary of the city.

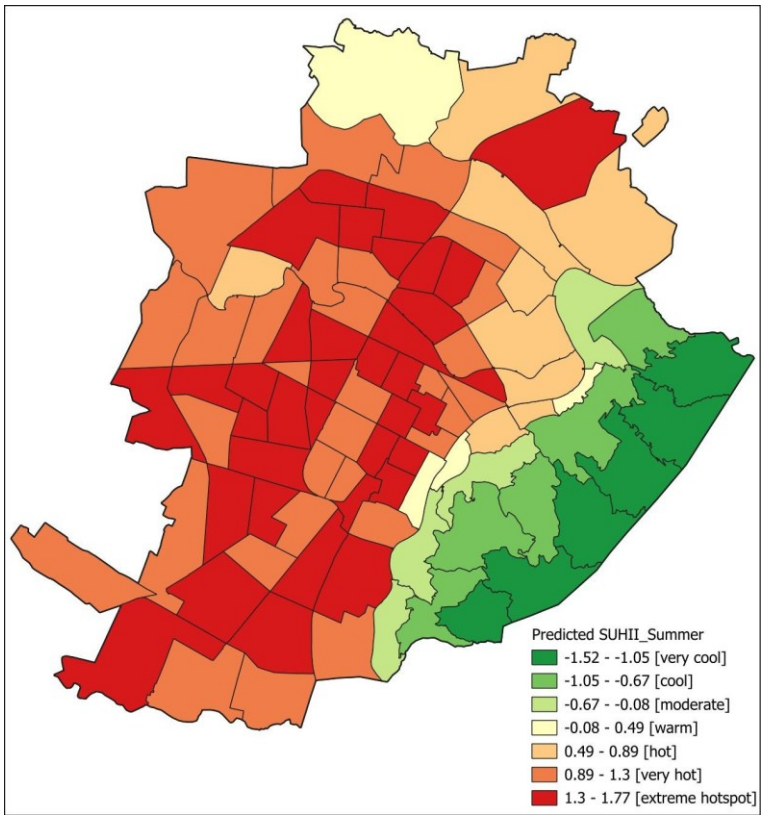


Figure 5.23. Area-weighted mean predicted SUHII by Zone Statistiche, summer 2025.

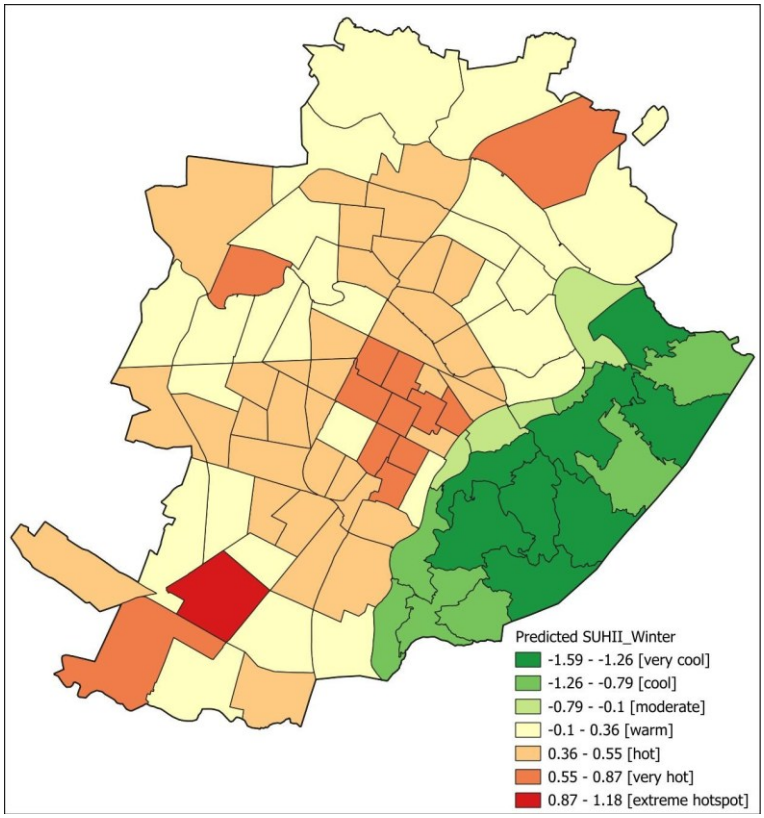


Figure 5.24. Area-weighted mean predicted SUHII by Zone Statistiche, winter 2025.

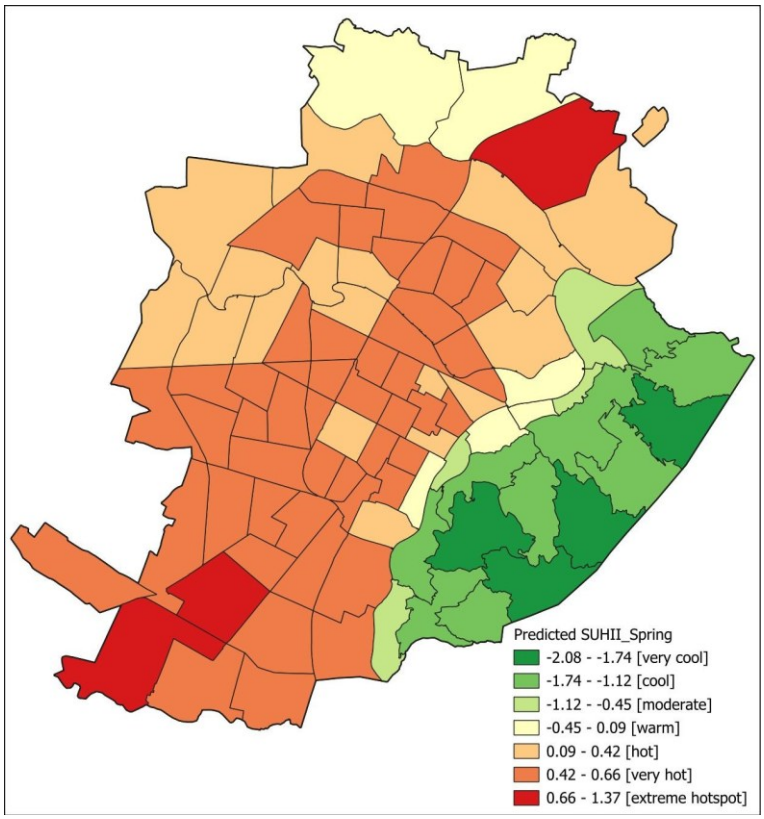


Figure 5.25. Area-weighted mean predicted SUHII by Zone Statistiche, spring 2025.

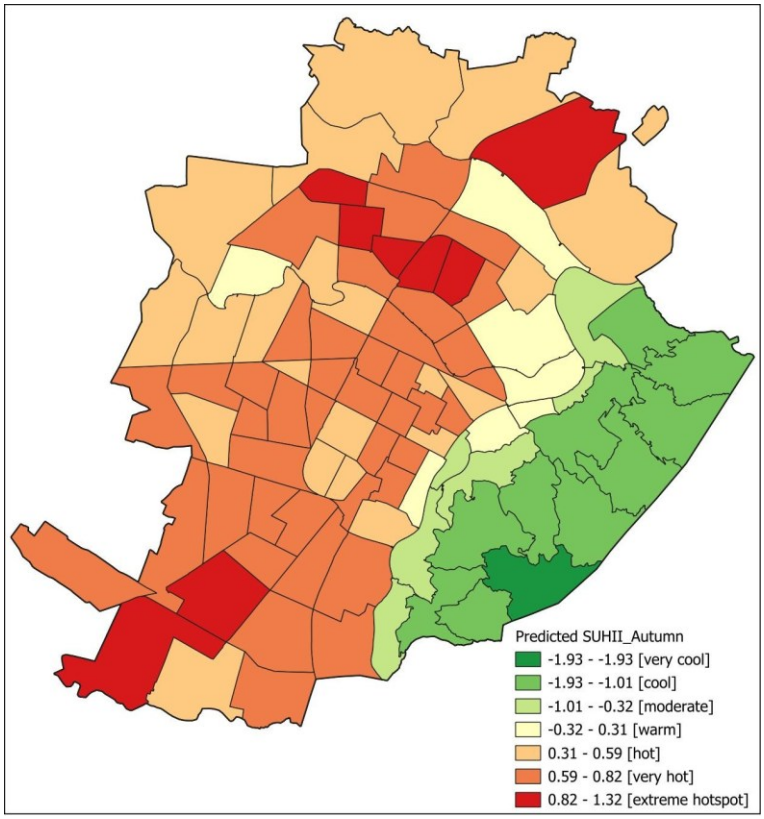


Figure 5.26. Area-weighted mean predicted SUHII by Zone Statistiche, autumn 2025.

5.10 Implications for the Next Stage

The Random Forest models and their SHAP attributions consistently identify three predictors that can be changed through physical intervention across all four seasons: NDVI, through vegetation cover; impervious surface fraction, through surface permeability; and albedo, through surface reflectance. Chapter 6 builds on these three predictors using the SHAP dependence-curve method of Li et al. (2025), which extracts quantitative mitigation thresholds and cooling estimates directly from the trained models, with no additional model runs on modified inputs.

One finding from this chapter is relevant to how the Chapter 6 results should be read. Albedo's Pearson correlation with SUHII is positive in winter, spring, and autumn (Tables 5.10, 5.12 and 5.14), and only weakly negative in summer, where

it is also flagged for a sign flip against its marginal sensitivity (Table 5.8). At the seasonal, grid-cell scale, cells with higher albedo tend to be modelled as warmer — a pattern driven by land-cover co-variation rather than by albedo's radiative properties. Chapter 6 interprets the albedo SHAP signal in light of this finding and draws on the physical literature where the model association and the surface physics differ.

The LCZ-based feasibility assessment, the zone-level comparison of mitigation potential against residual impact, and the C.I.R.C.A. environmental compensation framework are all developed in Chapter 6.

Chapter 6 — Mitigation Guidance and Environmental Compensation

6.1 From Modelled Drivers to Intervention Design

Chapter 5 established that NDVI, Impervious Surface Fraction (ISF), and Albedo recur as the dominant contributors to Surface Urban Heat Island intensity across all four seasons. NDVI alone accounts for over 40% of Gini feature importance in every season, with the largest model-implied sensitivity for both SUHII and SUHI. All three predictors are directly addressable through urban design decisions and carry a well-documented physical basis for their thermal effect. This chapter builds its intervention design around these three, setting aside the six morphological predictors in the retained set.

The methodology follows the SHAP dependence-curve framework developed by Li et al. (2025) for Hong Kong new towns.[57] Plotting each predictor's value against its SHAP contribution to predicted surface temperature yields four quantitative readings: a turning point, where the SHAP contribution changes direction; a warming peak, the maximum positive SHAP value and the feature value where it occurs; a saturation value, beyond which further adjustment yields diminishing returns; and a cooling peak, the SHAP value at saturation. Liu et al. (2025), applying a CatBoost-SHAP model across Local Climate Zone types in Shenzhen, found consistent turning-point behaviour for vegetated-cover and albedo predictors, supporting the generality of the dependence-curve approach beyond a single dataset.

[57]Li, Y., Zhang, H., Lin, Y., Ling, J., Xue, H., & Guo, P. (2025). Precise mitigation strategies for urban heat island effect in Hong Kong's new towns using automated machine learning. Sustainable Cities and Society, 125, 106350.

The chapter proceeds as follows. Section 6.2 presents SHAP feature importance plots and turning-point readings for the three actionable predictors, across all four seasons. Section 6.3 applies the Count of Warming Features (CWF) classification to map spatial urgency. Section 6.4 sets out the LCZ-based feasibility screening. Section 6.5 quantifies mitigation potential on feasible cells. Section 6.6 estimates the residual thermal burden on structurally constrained cells, defining the compensation requirement. Section 6.7 links that requirement to the C.I.R.C.A. environmental compensation framework. Section 6.8 gives zone-level recommendations. Section 6.9 consolidates all outputs into a spatial priority map.

6.2 SHAP Feature Importance and Seasonal Driver Profiles

Figures 6.1 and 6.2 show the SHAP beeswarm plots for summer 2025, for SUHI and SUHII respectively. Each point represents one grid cell. Its horizontal position shows the magnitude and direction of that predictor's SHAP contribution, and its colour encodes the cell's feature value on a blue-to-red scale. Three structural patterns emerge from these plots and hold across all four seasons.

First, NDVI produces the widest spread of any predictor, with its summer SHAP contribution running from roughly +4°C SUHI at the lowest NDVI values down to −8°C at the highest. This spread is wider in summer than in the other seasons, reflecting the vegetation cycle's amplifying effect on thermal contrasts during the peak growing season. Second, ISF ranks second in summer importance, and high-ISF cells generate positive SHAP contributions. Third, Albedo shows an asymmetric scatter in summer, with a cluster of strongly negative SHAP values at

very low Albedo — a confound addressed in Section 6.2.3. Equivalent beeswarm plots for Winter, Spring, and Autumn are provided in Appendix D.3.

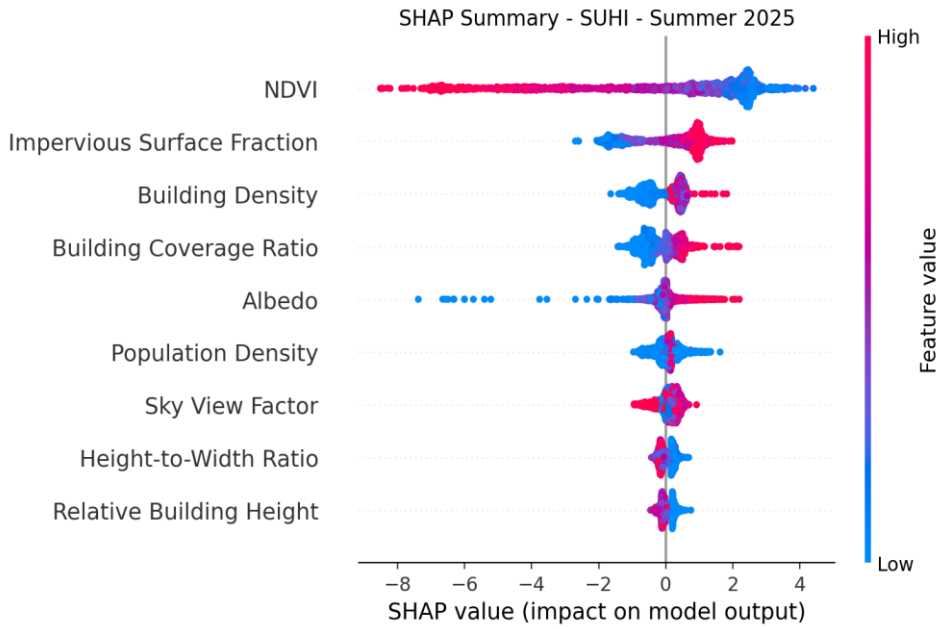


Figure 6.1. SHAP beeswarm — SUHI target, Summer 2025. Each point = one grid cell; horizontal = SHAP contribution (°C); colour = feature value (red high, blue low). Predictors ranked by mean |SHAP|.

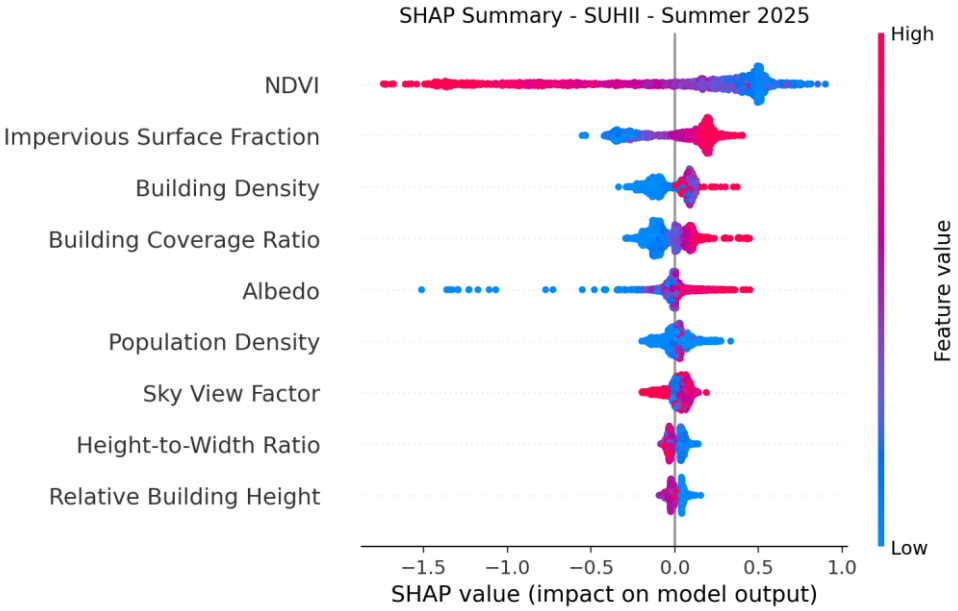


Figure 6.2. SHAP beeswarm — SUHII target, Summer 2025. The predictor ranking mirrors the SUHI beeswarm; axis scale is narrower because SUHII is a dimensionless z-score.

6.2.1 NDVI Dependence Curve

Figure 6.3 shows the SHAP dependence curve for NDVI against the SUHI target for summer 2025. The orange smoothed line through the scatter reveals a monotonic decline from left to right: cells below the turning point (NDVI = 0.507) contribute positively to SUHI, while cells crossing the turning point transition into the cooling regime. Low NDVI values (approximately 0.01) contributes to high warming temperatures of up to +2.458 °C. This indicates that there is a strong thermal influence in areas with almost no vegetation. Increasing NDVI to around 0.507 in summer can capture most of the potential cooling effect and is still a realistic greening target. A year-round NDVI above 0.574 would provide more consistent cooling benefits across seasons, as the seasonal threshold increases from 0.405 in spring to 0.574 in autumn.

The NDVI threshold changes from 0.405 in spring and 0.437 in winter to 0.507 in summer and 0.574 in autumn. A year-round target above 0.574 can provide cooling benefits.

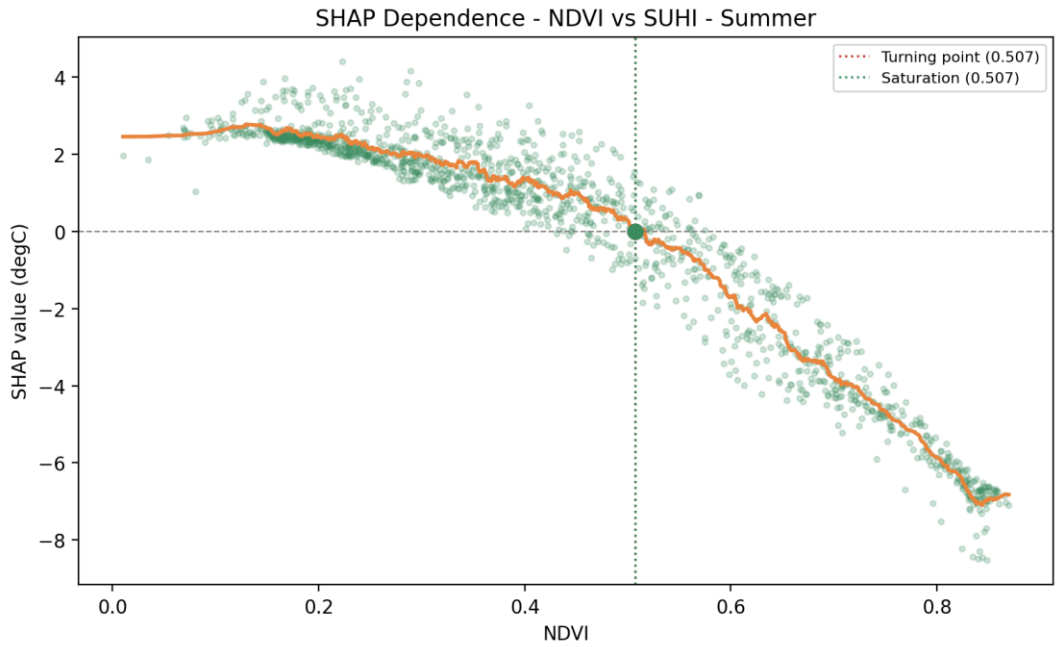


Figure 6.3. SHAP dependence curve — NDVI vs SUHI, Summer 2025. Red dotted line = turning point (0.507); grey dotted line = saturation value (0.507). Cells left of the line carry positive (warming) SHAP contributions; cells right carry negative (cooling) contributions.

6.2.2 Impervious Surface Fraction Dependence Curve

Figure 6.4 shows the ISF dependence curve for summer 2025. Two features are notable for planning. First, the turning point (0.504) and the saturation point (0.408) are distinct in summer — reducing ISF from above 0.504 to the saturation value of 0.408 captures the full available cooling benefit (−0.976 °C), without requiring the more radical transformation a very low ISF target would entail. Second, cells with ISF near zero already carry substantial negative SHAP values, confirming that de-paving and vegetation reinforce one another in the same direction. The ISF turning point is broadly stable across seasons (0.504 in summer, 0.565 in spring and autumn, 0.607 in winter), while the saturation value consistently falls in the 0.408–0.484 range, confirming that partial de-paving to the 40–50 % range is the robust cross-season target.

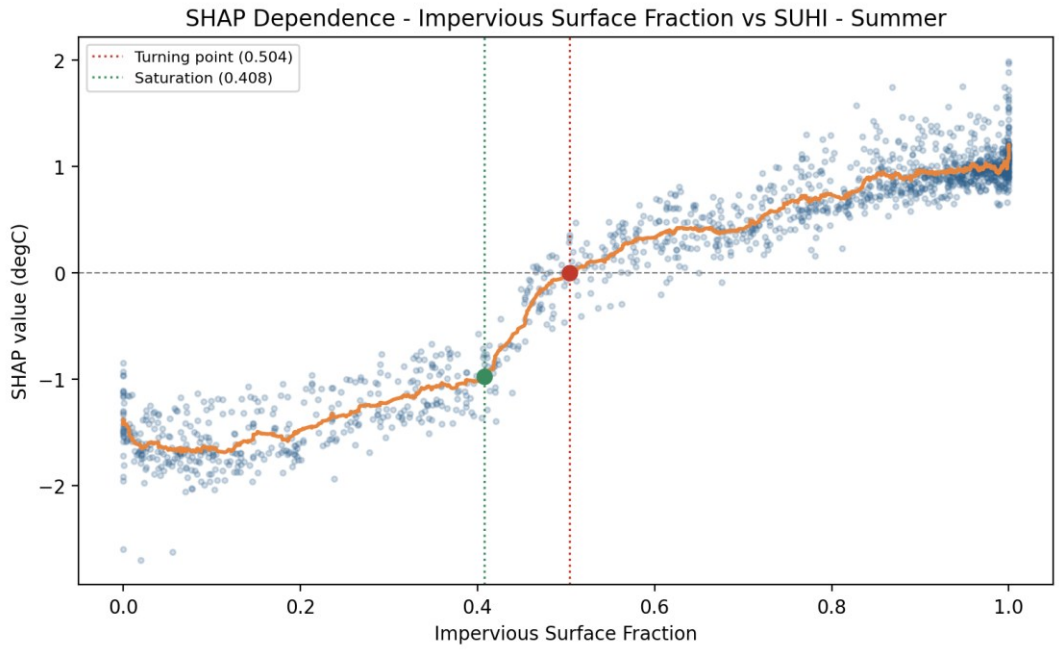


Figure 6.4. SHAP dependence curve — Impervious Surface Fraction vs SUHI, Summer 2025. Red dotted line = turning point (0.504); green dotted line = saturation value (0.408), where maximum cooling of −0.976 °C per cell is reached.

6.2.3 Albedo Dependence Curve and the Land-Cover Confound

Figure 6.5 shows the relationship between Albedo and SHAP values for summer 2025. Very low Albedo values (0.05–0.20) are associated with strongly negative SHAP values, reaching approximately −4 to −7 °C. The curve then changes rapidly around the turning point at 0.270 and reaches a nearly flat plateau, with slightly positive SHAP contributions above 0.35.

The strong negative SHAP values at low Albedo mainly occur in cells with high NDVI and low ISF, where vegetation cover is likely contributing to the observed cooling. The turning point at 0.270 and the small cooling effect near −0.007 °C should therefore be interpreted alongside these vegetation characteristics. Albedo can support cooling as a complementary measure, while its potential as

a primary intervention appears limited in this analysis. Further discussion of this limitation is provided in Chapter 7.

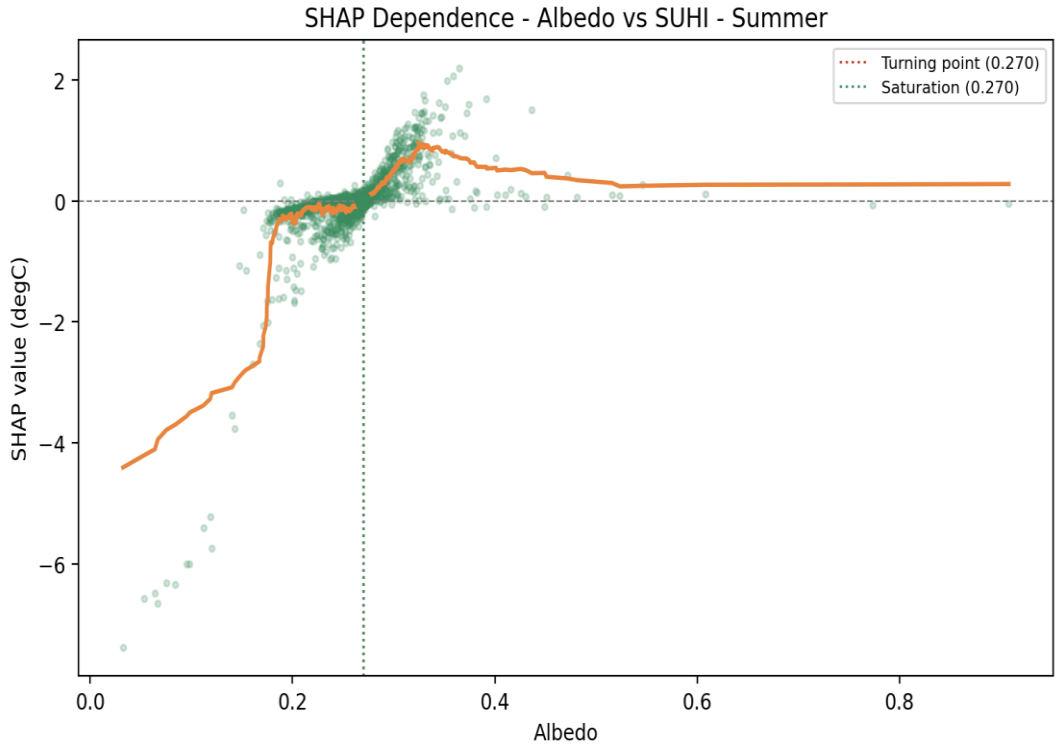


Figure 6.5. SHAP dependence curve — Albedo vs SUHI, Summer 2025. Turning point at Albedo = 0.270. The strongly negative SHAP values at very low Albedo reflect high-vegetation cells where Albedo and NDVI co-vary; the plateau above 0.35 reflects the high-ISF built core.

Season	Predictor	Turning Point	Warming Peak (°C)	Saturation Value	Cooling Peak (°C)
Winter	NDVI	0.437	+0.270	0.437	+0.004
Winter	ISF	0.607	+0.338	0.484	−0.235
Winter	Albedo	0.126	+0.240	0.126	−0.001
Spring	NDVI	0.405	+0.319	0.405	+0.016
Spring	ISF	0.565	+0.489	0.463	−0.374
Spring	Albedo	0.166	+0.405	0.166	−0.008
Summer	NDVI	0.507	+2.458	0.507	+0.001
Summer	ISF	0.504	+1.160	0.408	−0.976
Summer	Albedo	0.270	+0.281	0.270	−0.007
Autumn	NDVI	0.574	+0.753	0.577	−0.006
Autumn	ISF	0.565	+0.686	0.463	−0.437
Autumn	Albedo	0.208	+0.467	0.208	−0.004

Table 6.1. SHAP dependence-curve readings for all three actionable predictors across all four seasons (SUHI target). Turning points are in raw predictor scale; warming and cooling peaks in °C SUHI. Source: nine-predictor Random Forest models, Chapter 5.

6.3 Spatial Priority Classification: Count of Warming Features

The Count of Warming Features (CWF) classification assigns each grid cell a score equal to the number of the nine retained predictors whose SHAP contribution is positive for that cell in a given season.58

Combining each cell's SUHI with its CWF score creates four priority classes: Most Emergency, Remain Concerned, Less Emergency, and Well-Built, based on

whether SUHI is above or below the seasonal mean and CWF is above or below the median. This classification highlights areas where high SUHI occurs together with several warming factors, indicating a higher priority for intervention. Cells with fewer warming factors may require more targeted measures focused on the specific predictors contributing to their heat levels.

Figure 6.6 shows the CWF class distribution for summer 2025. Most Emergency cells dominate at 848 (56.5 % of the subsample), confirming that the majority of above-average SUHI cells in summer also carry multiple simultaneous warming predictor contributions. Well-Built cells (455) concentrate in the greenway belt and lower-density peripheral zones. The spatial distribution across the Turin municipal boundary is shown in Figure 6.7, where deep red cells represent primary targets for intervention.

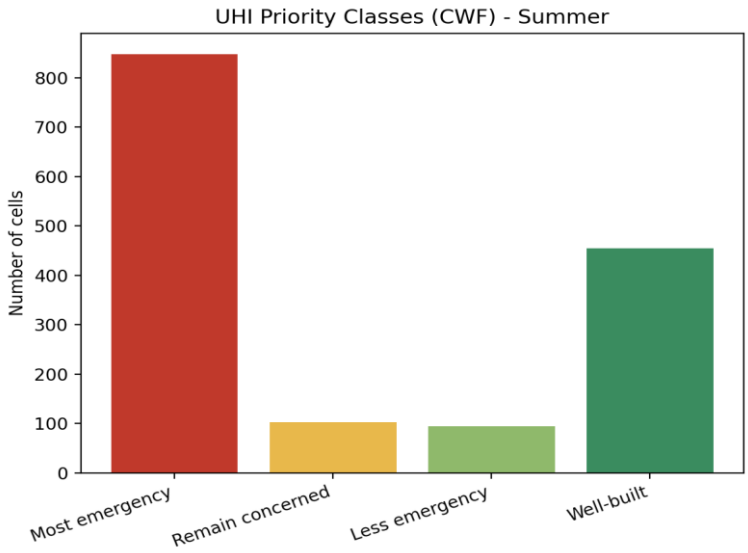


Figure 6.6. CWF priority class distribution — Summer 2025. Most Emergency cells (848) carry above-average SUHI combined with above-median count of warming predictors. Well-Built cells (455) carry below-average SUHI and below-median CWF.

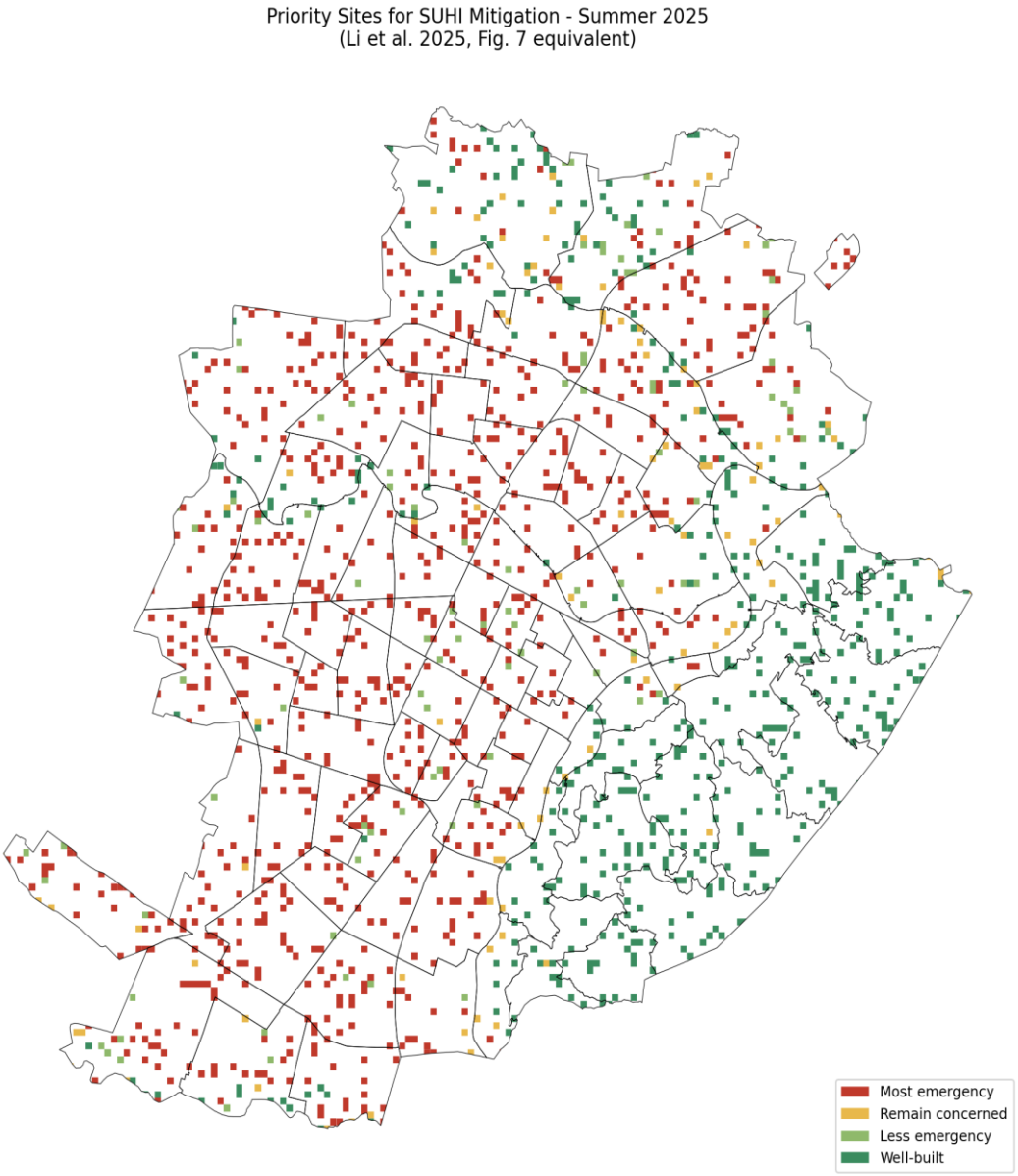


Figure 6.7. Spatial distribution of CWF priority classes across Turin — Summer 2025 (Li et al. 2025, Fig. 7 equivalent). Deep red = Most Emergency. Green = Well-Built. The consolidated urban core and mid-rise residential grid carry the highest concentration of Most Emergency cells.

Season	Most Emergency	Remain Concerned	Less Emergency	Well-Built
Winter	508	189	369	434
Spring	787	178	152	383
Summer	848	103	94	455

Autumn	763	160	168	409

Table 6.2. CWF priority class distribution across all four seasons. Most Emergency cells peak in summer; Well-Built cells are proportionally higher in winter when absolute SUHI contrasts are lower.

6.4 LCZ-Based Feasibility Screening

Whether a Most Emergency cell can receive on-site physical intervention depends on its built form. The feasibility classification assigns cells to one of three categories. Cells in LCZ 1 (Compact High-Rise) or LCZ 2 (Compact Mid-Rise) with ISF at or above the seasonal median are classified Infeasible — structural density leaves insufficient permeable or plantable surface area for meaningful NDVI enhancement or ISF reduction. Cells in all remaining built LCZ types (3 through 10) meeting the ISF threshold are classified Feasible. Non-built LCZ types (11, 12, 14, 17) are Excluded from both mitigation and compensation accounting.

For summer 2025, the classification yields 441 Feasible cells, 284 Infeasible cells, 326 non-built Excluded cells, and 449 built low-ISF Excluded cells. The Infeasible cells concentrate in the centro storico, Crocetta, San Salvario, and the dense residential grid around Porta Nuova. The classification is re-applied season by season using the seasonal ISF median, so a cell's classification can shift across seasons.

6.5 Mitigation Potential on Feasible Cells

The mitigation potential for each predictor is estimated by identifying non-compliant cells — those whose predictor value places them on the warming side of the seasonal turning point — and calculating the SHAP-implied cooling from shifting each cell's predictor to the turning point value. Figure 6.8 shows the city-wide total for summer 2025 across all sampled cells.

NDVI delivers the largest total benefit at −1,793.3 °C·cells across 957 non-compliant cells (mean −1.874 °C per cell). ISF follows at −810.1 °C·cells across 998 cells (mean −0.812 °C). Albedo is substantially smaller at −204.9 °C·cells (mean −0.431 °C). The dominance of NDVI in summer — carrying more than twice the ISF potential — is driven by the size of the summer warming peak (+2.458 °C) and the large fraction of cells whose NDVI falls below the turning point. NDVI is also the most seasonally robust lever: the winter per-cell benefit (mean −0.373 °C) reaches approximately 40 % of the summer value, meaning that greening interventions retain partial benefit throughout the year.

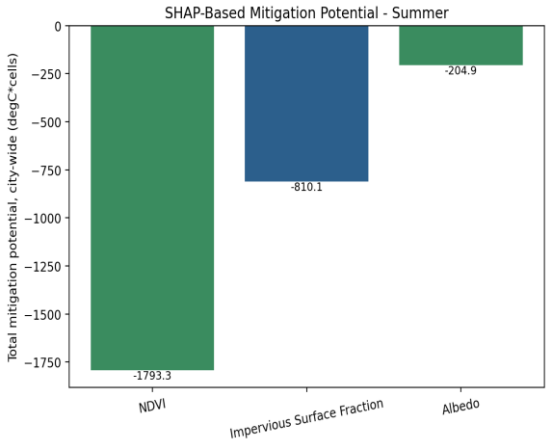


Figure 6.8. City-wide mitigation potential — Summer 2025. Total SHAP-implied cooling (°C·cells) achievable if all non-compliant cells are brought to their seasonal turning point. Negative values indicate cooling. NDVI: −1,793.3; ISF: −810.1; Albedo: −204.9.

Season	Predictor	Non-Compliant Cells	Mean Benefit (°C/cell)	Mean Benefit (°C)
Winter	NDVI	425	−0.373	-0.373
Winter	ISF	438	−0.174	-0.174
Winter	Albedo	292	−0.332	-0.332

Season	Predictor	Non-Compliant Cells	Mean Benefit (°C/cell)	Mean Benefit (°C)
Spring	NDVI	421	−0.571	-0.571
Spring	ISF	438	−0.312	-0.313
Spring	Albedo	292	−0.468	-0.468
Summer	NDVI	421	−1.874	-1.874
Summer	ISF	438	−0.812	-0.812
Summer	Albedo	292	−0.431	-0.431
Autumn	NDVI	395	−0.705	-0.705
Autumn	ISF	438	−0.476	-0.476
Autumn	Albedo	292	−0.285	-0.285

Table 6.3. SHAP-based mitigation potential on LCZ-feasible cells, by predictor and season (SUHI target). Non-Compliant Cells = feasible cells with predictor on the warming side of the seasonal turning point.

6.6 Residual Thermal Burden and the Compensation Requirement

The residual thermal burden is the SHAP-implied cooling that would accrue if infeasible cells were brought to the predictor turning point but which cannot be realised on-site given the structural classification of those cells. This quantity defines the minimum compensation requirement: the cooling that must be delivered through interventions elsewhere to offset the heat that structurally infeasible cells continue to generate. Figure 6.9 shows the split between mitigation potential (feasible cells, green) and residual impact (infeasible cells, red) for summer 2025.

For NDVI, 800 °C·cells are achievable through on-site mitigation in feasible cells, while 652 °C·cells represent the residual burden in infeasible cells. For ISF, the split is 409 °C·cells (feasible) versus 281 °C·cells (infeasible). Albedo carries a much smaller residual (1 °C·cell) because the ISF-based infeasibility criterion is less restrictive for cool-roof or reflective-surface interventions.

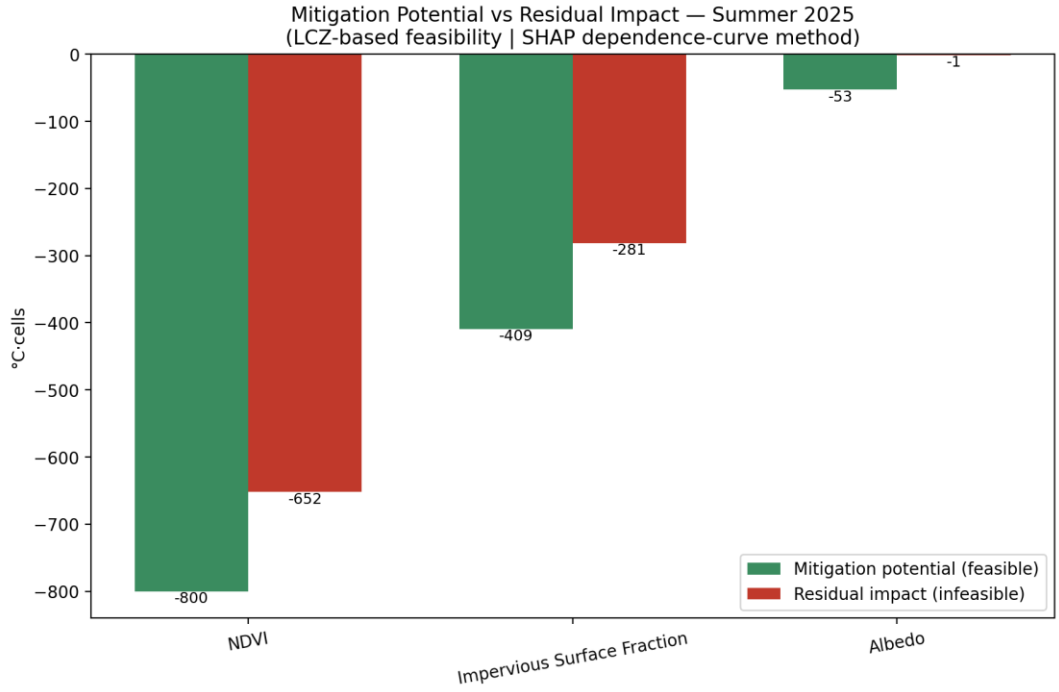


Figure 6.9. Mitigation potential (feasible cells, green) vs residual impact requiring compensation (infeasible cells, red) — Summer 2025. NDVI: feasible −800 / residual −652 °C·cells. ISF: feasible −409 / residual −281. Albedo: feasible −53 / residual −1.

Predictor	Population	Cells (all seasons)	Total (°C·cells)	Mean Benefit (°C/cell)
NDVI	Mitigation (feasible)	1,662	−1,557.8	−0.937
NDVI	Residual / Compensation	1,147	−1,268.5	−1.106
ISF	Mitigation (feasible)	1,761	−866.8	−0.492

ISF	Residual / Compensation	1,147	−617.7	−0.539
Albedo	Mitigation (feasible)	883	−356.9	−0.404
Albedo	Residual / Compensation	154	−33.9	−0.220

Table 6.4. Annual aggregated mitigation potential (feasible cells) and residual impact (infeasible cells), summed across all four seasons. Negative values indicate cooling potential (SUHI target).

The mean per-cell residual benefit for NDVI (−1.106 °C) slightly exceeds the mean mitigation benefit (−0.937 °C), indicating that the most thermally intense cells — those that tend toward infeasibility — carry above-average SHAP-implied cooling potential.

6.7 Environmental Compensation and the C.I.R.C.A. Framework

Environmental compensation within the Italian regulatory context is defined by ISPRA (2015) as interventions carried out to offset non-mitigatable residual environmental impacts — actions not tied to the transforming land use, but designed to restore or improve equivalent environmental conditions elsewhere.[59]

Within Turin's metropolitan planning framework, this principle is operationalised through the C.I.R.C.A. catalogue (Catalogo Interventi di Riqualificazione e Compensazione Ambientale), introduced in the 2023 General Territorial Metropolitan Plan.[60]

[59]ISPRA (2015)

[60]Barbieri, Bottero, Caprioli, Cassatella, Datola, D'Errico, Erbetta, & Gottero (2026)

The Abaco delle Compensazioni Ambientali (Cassatella et al., 2025) organises forty-four Nature-Based Solutions into eight intervention categories.[61]

Three categories directly address the SHAP-identified residual burden:

Category B — Climate Change and Urban Heat Islands:
Category C — Forestation:
Category G — Green Infrastructure:

Dominant Driver	Annual Residual (°C.cells)	C.I.R.C.A. Objective	Abaco Category	Key Intervention Types
NDVI	−1,268.5	Resilience / Green Infra.	B, C, G	Urban tree planting, green corridors, green roofs (intensive), urban meadows
ISF	−617.7	Environmental quality	B, G	Permeable paving, de-paving + renaturalisation, sustainable urban drainage (SUDs)
Albedo	−33.9	Environmental quality	B	Cool roofs (extensive green), reflective paving, albedo-enhancing surface treatments

Table 6.5. Linkage between SHAP-identified residual thermal burdens and C.I.R.C.A. / Abaco intervention categories (Cassatella et al., 2025).

6.8 Zone-Specific Recommendations

Figures 6.10 to 6.12 provide three complementary spatial views of the summer dominant-driver analysis. Figure 6.10 maps the dominant actionable driver in Most Emergency cells by statistical zone. Figure 6.11 maps the dominant predictor across all above-average SUHI cells (the broader warming-driver picture). Figure 6.12 maps the dominant predictor in below-average SUHI cells, identifying where current cooling conditions might be extended.

[61]Cassatella, C., Ferrari, S., Bottero, M., Caprioli, C., Barbieri, S., & Erbetta, A. (2025). Abaco delle Compensazioni Ambientali. Politecnico di Torino – DIST / Città Metropolitana di Torino. DOI: 10.5281/zenodo.15462901.

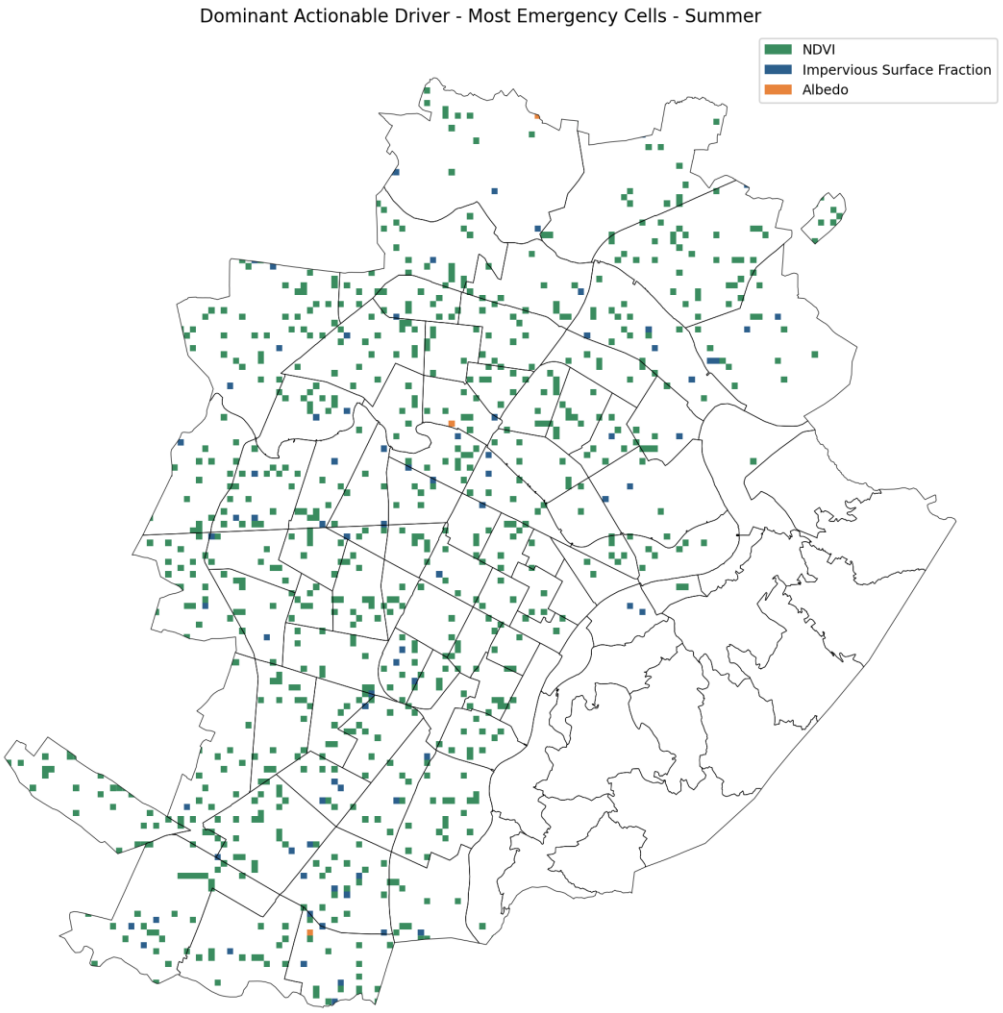


Figure 6.10. Dominant actionable driver in Most Emergency cells — Summer 2025. Green = NDVI; Blue = ISF; Orange = Albedo. NDVI dominates virtually all priority zones in summer; ISF appears as primary driver in a subset of inner-peripheral zones.

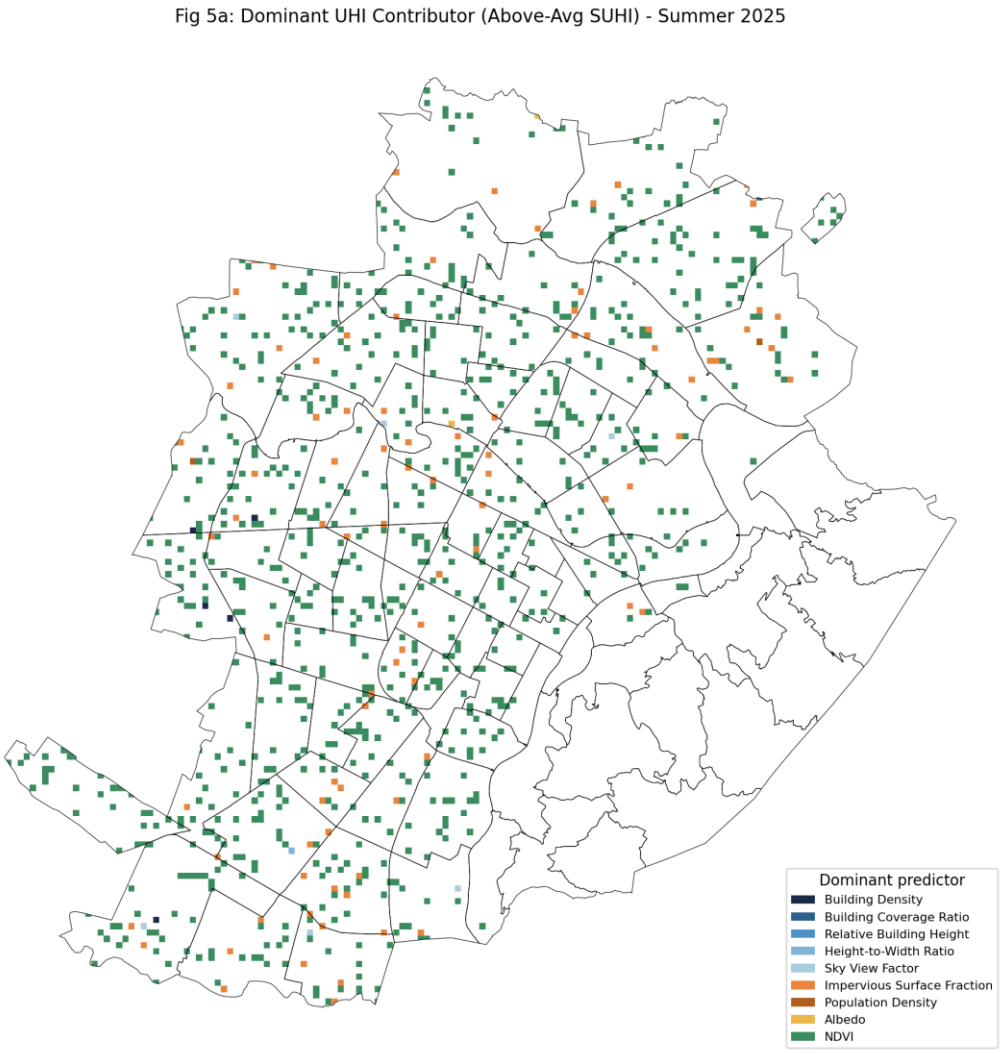


Figure 6.11. Dominant predictor in above-average SUHI cells (all 9 predictors) — Summer 2025. NDVI (green) accounts for the majority of dominant-warming cells across the city.

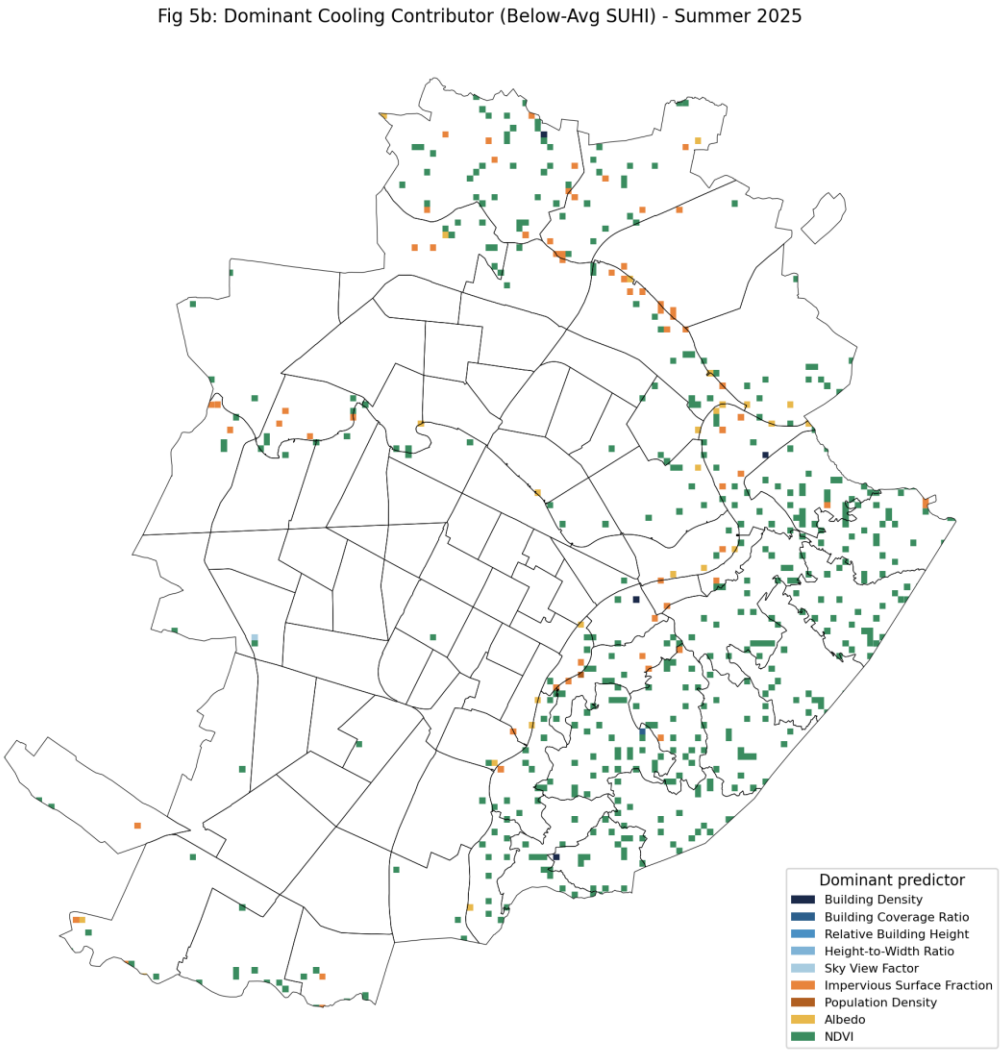


Figure 6.12. Dominant predictor in below-average SUHI cells — Summer 2025. NDVI dominates the cooling side, concentrated in the northern and eastern green belt and along the river corridor.

In summer, NDVI is the exclusive primary actionable driver across all zones recording 100 % Most Emergency status, confirming the summer warming peak (+2.458 °C) as the strongest single-predictor signal. In winter, the driver pattern shifts: Albedo becomes the primary driver in Villaggio Snia (88 % Most Emergency), Tre Tetti — Bellezia (72 %), and Vallette — Saffarona (65 %), where ISF and NDVI move to secondary positions. This seasonal rotation provides the

basis for season-differentiated planning: summer priorities call for NDVI-targeted greening, while winter priorities indicate reflective surface interventions consistent with Abaco Category B.

Zone	% Emergency	Emergency Cells	Primary Driver	Turning Point	Warming Peak (°C)	Fallback
Barriera di Orbassano	100%	18	NDVI	0.507	+2.458	—
Borgata Cenisia	100%	8	NDVI	0.507	+2.458	—
Borgo San Paolo	100%	13	NDVI	0.507	+2.458	—
Borgo San Secondo — Staz. P. Nuova	100%	8	NDVI	0.507	+2.458	—
Borgata Monterosa	100%	10	NDVI	0.507	+2.458	—
Borgata Vittoria	100%	9	NDVI	0.507	+2.458	—
Borgata Montebianco	100%	9	NDVI	0.507	+2.458	—
Santa Rita Da Cascia	100%	8	NDVI	0.507	+2.458	—
Polo Nord	100%	12	NDVI	0.507	+2.458	ISF
Pozzo Strada	100%	8	NDVI	0.507	+2.458	ISF
Ospizio di Carità	100%	8	NDVI	0.507	+2.458	ISF

Table 6.6. Zone-specific recommendations for the highest-urgency summer zones (>30 % Most Emergency cells). Turning points and warming peaks from Table 6.1.

6.9 Spatial Priority Mapping

Figure 6.13 combines the seasonal mitigation and compensation results into a planning-oriented spatial output for summer 2025. The dual choropleth maps show mitigation potential in feasible cells and residual compensation requirements in infeasible cells at the statistical-zone level. The mitigation map indicates the highest potential in the mid-ring and inner peripheral zones. The compensation map shows the highest requirements in the dense historic core and compact mid-rise districts.

.

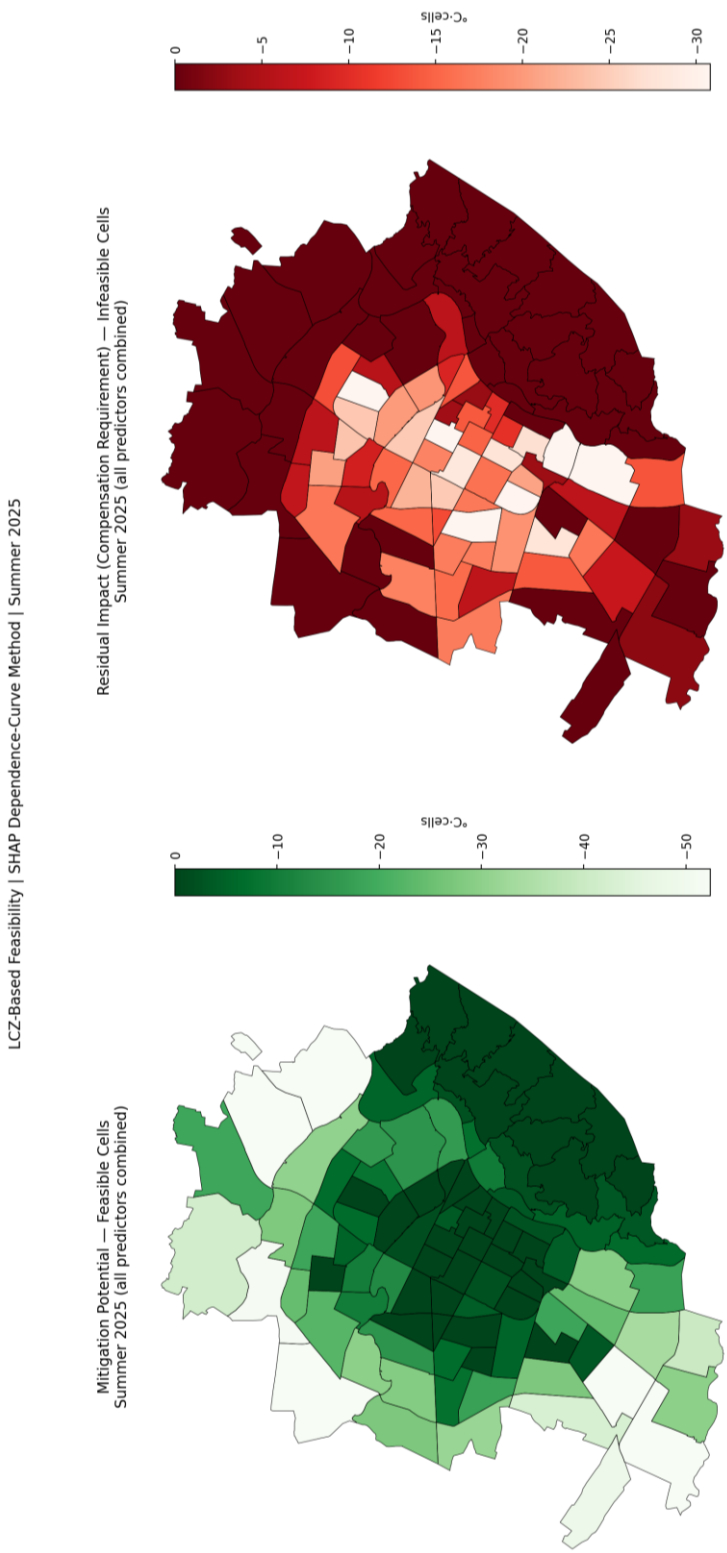


Figure 6.13. Zone-level mitigation potential (green, feasible cells) and residual compensation requirement (red, infeasible cells) — Summer 2025. Deeper green = more on-site cooling available. Deeper red = larger compensation burden. The spatial inversion reflects the LCZ-based feasibility split: zones with the greatest thermal need carry the least feasible intervention space.

Statistical Zone	*Total Cells*	*Mitigation (°C)*	*Residual (°C)*
Borgo San Paolo	52	+0.000	-1.473
Corso Dante — Ponte Isabella	60	-0.132	-1.203
Borgata Monterosa	42	+0.000	-1.683
Municipio	37	+0.000	-1.800
Molinette — Vecchia Fiat — Stazione Smistamento	72	-0.872	-0.883
Crocetta	44	+0.000	-1.441
Comandi Militari — Stazione Porta Susa	36	+0.000	-1.644
Santa Rita Da Cascia	32	+0.000	-1.738
Borgo San Secondo — Stazione Porta Nuova	32	+0.000	-1.716
Piazza Nizza — (Borgo San Salvario)	32	+0.000	-1.712
Borgo Dora — Valdocco	42	-0.090	-1.252
Mercato del Bestiame — Aiuola Martini	32	+0.000	-1.587
Borgata Montebianco	38	-0.434	-1.295
Borgata Vittoria	34	-0.256	-1.424
Borgo San Donato	28	+0.000	-1.639

Table 6.7. Top 15 statistical zones by annual cumulative residual thermal burden (°C·cells), all four seasons combined. Zones with zero mitigation have no sampled feasible cells in the analysis.

Nine of the fifteen highest-burden zones record zero on-site mitigation potential, indicating that their sampled cells fall entirely in the infeasible or excluded categories. These zones — including Borgo San Paolo, Borgata Monterosa, and Municipio — are the clearest candidates for designation as C.I.R.C.A. compensation-receiving areas, where greening, de-paving, and reflective surface programmes should be preferentially directed through the next CIRCA notification cycle.

Chapter 7 — Discussion and Conclusions

7.1 Summary of Findings

This thesis analysed the seasonal drivers of Surface Urban Heat Island intensity in the Metropolitan City of Turin and translated those findings into a spatially differentiated, evidence-grounded mitigation and compensation framework. The analysis integrated four-season Landsat remote sensing, Local Climate Zone classification, nine-predictor Random Forest modelling with SHAP interpretation, and SHAP dependence-curve-based intervention guidance linked to the C.I.R.C.A. environmental compensation catalogue.

Chapter 5 established NDVI, Impervious Surface Fraction, and Albedo as consistent dominant contributors to both SUHII and SUHI across all four seasons. NDVI contributed more than 40% of the Gini feature importance in all seasons. Mean |SHAP| values of NDVI range from 2.537 °C in summer, 0.743 °C in spring, 0.492 °C in winter, and 0.372 °C in autumn. The nine-predictor Random Forest models achieved test-set R^2 values of 0.902 in summer, 0.809 in autumn, 0.807 in spring, and 0.701 in winter.

Chapter 6 applied the SHAP dependence-curve method of Li et al. (2025) to derive season-specific intervention thresholds for NDVI (turning points 0.405–0.574), ISF (0.504–0.607), and Albedo (0.126–0.270). On feasible cells, NDVI enhancement delivered the largest modelled cooling potential in every season, reaching −1.874 °C per cell in summer. The LCZ-based feasibility split revealed that structurally infeasible cells carry an annual NDVI cooling deficit of −1,268.5 °C·cells, requiring an external compensation programme. Nine of the fifteen highest-burden statistical zones recorded zero on-site mitigation capacity, with

Borgo San Paolo, Borgata Monterosa, and Municipio carrying the largest annual compensation burdens.[62]

The C.I.R.C.A. catalogue[63] and Abaco delle Compensazioni Ambientali (Cassatella et al., 2025) provide the institutional and technical infrastructure for directing those burdens toward greening, de-paving, and reflective surface interventions in registered compensation-receiving areas.[64]

7.2 Methodological Contributions

Three methodological contributions distinguish this study from a conventional heat-mitigation scenario study. The first is the use of SHAP dependence curves — rather than fixed-increment simulations or generic planning targets — to derive intervention thresholds that are specific to Turin's own dataset and trained models. The turning points in Table 6.1 emerge from the fitted statistical relationships between Turin's urban fabric and its seasonal SUHI patterns, grounding every intervention threshold in the observed distribution of conditions among comparable cells.

The second contribution is the use of SHAP values to define the spatial feasibility boundary in a way that is internally consistent with the model used to assess mitigation potential. The same SHAP framework identifies which predictor is the

dominant driver in each zone and quantifies the cooling that would follow from addressing that driver, eliminating the mismatch between the model's implicit assumptions and the scenario's explicit intervention assumptions.

The third contribution is the direct linkage between the SHAP-derived residual thermal burden and the C.I.R.C.A. institutional framework. The connection to a specific, legally grounded, metropolitan-scale compensation catalogue, with the quantification of a zone-by-zone compensation requirement readable against C.I.R.C.A.-registered receiving areas, represents a step toward making modelling outputs actionable within an existing planning workflow.

7.3 Limitations

One key limitation is the observational nature of the Albedo–SUHI relationship at the grid-cell scale. The positive correlation between Albedo and surface temperature, consistent across winter, spring, and autumn in both the Pearson correlation and the SHAP dependence curve, reflects land-cover co-variation rather than the physical radiative effect of surface reflectance: high-Albedo grid cells in Turin tend to be high-ISF, low-NDVI cells. A correlational Random Forest cannot fully separate the reflectivity signal from the land-cover configuration signal at this spatial scale. Physical-process models or experimental interventions would be required to quantify the Albedo cooling effect precisely.

A second limitation is the simplified LCZ-based feasibility criterion. The classification used here — LCZ 1/2 combined with above-median ISF — is a single-variable proxy for the many site-specific constraints that govern real intervention feasibility: heritage listing, structural load capacity of existing roofs,

underground utility conflicts, and land tenure are all genuinely relevant and entirely absent from the present classification.

As the 100 × 100 m grid captures patterns at the city-block scale, while details of individual buildings, courtyards, and street layouts are not represented, this forms the third limitation. The Landsat-derived SUHI values also reflect the sensor's spatial averaging. In addition, the four seasonal images represent specific thermal conditions for each season rather than continuous conditions throughout the year.

7.4 Planning Guidance

For the Turin Metropolitan Authority, the outputs offer three levels of actionable guidance. At the city-wide level, the CWF priority maps of Chapter 6 provide a seasonally stratified spatial layer that can be overlaid on any existing thermal vulnerability assessment to identify cells combining high SUHI, multiple warming predictors, and feasible on-site intervention.

At the zone level, the dominant-driver analysis of Section 6.8 allows the C.I.R.C.A. compensation programme to be steered by each zone's specific predictor deficit. Summer NDVI-deficient zones call for canopy planting and green corridor creation (Abaco Categories B and C). Winter Albedo-driven zones — Villaggio Snia, Tre Tetti, Vallette — benefit from reflective surface measures as primary interventions (Abaco Category B), complemented by greening as the thermal season lengthens. ISF-driven fallback zones benefit from permeable paving and de-paving programmes (Abaco Category G and the Sustainable Urban Drainage Systems subsection).

At the compensation planning level, the fifteen-zone residual burden ranking of Table 6.7 offers a prioritised list of zones that should be designated as primary C.I.R.C.A. compensation recipients. The nine zones with zero on-site mitigation capacity represent the clearest cases for off-site compensation, where the C.I.R.C.A. catalogue's receiving-area designation mechanism should be applied.

7.5 Future Work

7.5.1 SHAP-Based Evaluation of Compensatory Measures

The most immediate extension is the application of the same SHAP dependence-curve method to evaluate the cooling performance of specific C.I.R.C.A. compensation interventions before deployment. A SHAP-based evaluation would convert tree planting to an estimated NDVI increment, de-paving to an ISF reduction, and pass that predictor shift through the trained Random Forest to generate a predicted SUHI change. The turning points and saturation values established in Chapter 6 would then allow each proposed measure to be evaluated against the data-derived threshold, providing a quantitative link between intervention type, predicted thermal benefit, and the residual burden it is designed to offset.

A related development would be the formal derivation of a thermal offset ratio — cooling achievable per unit area of each intervention type, expressed in °C·m² — that could serve as a standardised currency for compensation agreements. Barbieri et al. (2025) identify the absence of such standardisation

as a key practical obstacle to effective compensation design in the Italian context.[65]

7.5.2 Expanded Feasibility Classification

The LCZ-based feasibility criterion of Chapter 6 would be substantially strengthened by incorporating heritage protection boundaries, structural assessment data for roof-based interventions, land-ownership registers distinguishing public from private plots, and utility conflict maps for sub-surface installations. A staged building-classification approach would provide a more precise and planning-defensible delineation of the infeasible stock, reducing both the over-classification of genuinely accessible sites and the under-classification of sites where structural constraints are more severe than the LCZ type alone implies.

7.5.3 Resolving the Albedo Confound

The Albedo confound identified in Section 6.2.3 merits a dedicated investigation using spatial causal-inference methods or physically based microclimate modelling capable of separating the radiative cooling mechanism from the land-cover co-variation signal. Spatial Durbin models or geographically weighted regression with explicit ISF controls would test whether the positive Albedo–SUHI association persists once ISF spatial autocorrelation is accounted for. Physical modelling approaches — mesoscale models such as WRF-UCM or street-scale computational fluid dynamics — would provide direct estimates of the surface energy balance change attributable to reflectivity improvements,

uncoupled from the observational correlations that the Random Forest inherits from the satellite record.

7.5.4 Multi-Year Temporal Generalisation

The four 2025 seasonal snapshots represent a single annual cycle. Extending the analysis across multiple years would test the stability of the turning points, the consistency of the zone-level ranking, and the seasonal reproducibility of the NDVI and ISF cooling potentials. In a warming climate, NDVI turning points may shift upward as baseline summer temperatures increase; tracking that shift over time would allow the intervention thresholds to be updated dynamically. A multi-year analysis would also enable construction of a thermal vulnerability index based on persistence and trend, providing a more robust basis for long-term capital investment decisions.

7.6 Concluding Remarks

This thesis has shown that a nine-predictor Random Forest model, trained on seasonal Landsat and morphological data and interpreted through SHAP dependence curves, can support a complete, spatially explicit, and institutionally grounded urban heat mitigation and compensation framework for the Metropolitan City of Turin. The framework moves from predictor identification through seasonal intervention thresholds to zone-level driver attribution, feasibility classification, and a quantified compensation requirement — all derived from a single coherent modelling pipeline.

The findings confirm NDVI, Impervious Surface Fraction, and Albedo as the three levers through which Turin's urban fabric most directly shapes its seasonal SUHI pattern, and the SHAP-derived turning points (NDVI 0.405–0.574 across seasons,

ISF 0.504–0.607) provide data-specific, reproducible targets for greening and de-paving investment. The spatial concentration of residual thermal burden in nine structurally infeasible zones maps directly onto the geographic coverage of the C.I.R.C.A. catalogue's first compensation cycle, providing a quantified spatial demand that the catalogue's receiving-area mechanism is designed to address.

The broader contribution lies in demonstrating how the same interpretable machine learning tools that diagnose the problem can be extended to guide the solution, and how those quantitative outputs can be coupled to an existing institutional planning framework. The standardisation of thermal offset ratios and the SHAP-based evaluation of specific compensatory measures — identified as the most immediate next step — would complete that link and allow the framework to function as a fully operational decision-support tool for Turin's environmental compensation programme.

References

Arnfield, A. (2003). Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology, 23*(1), 1-26.

ARPA Piemonte. (n.d.). Retrieved from https://www.arpa.piemonte.it/rischi_naturali/snippets_arpa_graphs/map_meteoweb/

Avdan, U., & Jovanovska, G. (2016). Algorithm for automated mapping of land surface temperature using LANDSAT 8 satellite data. *Journal of Sensors, 2016*, 1480307.

Barbieri, S., Bottero, M., Caprioli, C., Cassatella, C., & Datola, G. (2026). Guiding and Supporting Environmental Compensations. A Tool for Identifying, Estimating and Disseminating Best Offset Measures in Italian Environmental Impact Assessment and Strategic Environmental Assessment Processes. *Landscapes Across the Mediterranean (CrossMED), Reggio Calabria*, (pp. 139-144).

Bechtel, B., Alexander, P., Beck, C., Böhner, J., Brousse, O., & Ching, J. (2019). Generating WUDAPT Level 0 data – Current status of production and evaluation. *Urban Climate, 27*, 24-45.

Bowler, D., Buyung-Ali, L., Knight, T., & Pullin, A. (2010). Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning, 97*(3), 147-155.

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

Buccolieri, R., Esposito, A., Pappaccogli, G., Grulois, M., Donateo, A., Santiago, J., . . . Salizzoni, P. (2022). Local Climate Zones (LCZs) and Urban Morphological Parameters Using GIS: An Application to Italian Cities. *Environmental Sciences Proceedings 19, 15 (5th International Electronic Conference on Atmospheric Sciences)*, (pp. 1-6).

Clinton, N., & Gong, P. (2013). MODIS detected surface urban heat islands and sinks: Global locations and controls. *Remote Sensing of Environment, 134*, 294-304.

Demuzere, M., Bechtel, B., Middel, A., & Mills, G. (2020). Mapping Europe into local climate zones. *PLOS ONE, 14*(4), e0214474.

Demuzere, M., Kittner, J., & Bechtel, B. (2022). A global map of local climate zones to support earth system modelling and urban-scale environmental science. *Earth System Science Data, 14*(8), 3835-3873.

Ebrazi, R. (2026). *Integrated Modeling of the Surface Urban Heat Island Using Landsat Remote Sensing and GIS Tools: A Comparative Assessment of Wrocław and Turin.* Turin, Italy: Politecnico di Torino, Master's Thesis, Planning for the Global Urban Agenda.

Ellena, M., Melis, G., Zengarini, N., Di Gangi, E., Ricciardi, G., Mercogliano, P., & Costa, G. (2023). Micro-Scale UHI Risk Assessment on the Heat–Health Nexus within Cities by Looking at Socio-Economic Factors and Built Environment Characteristics: The Turin Case Study (Italy). *Urban Climate, 49*, 101514.

Ermida, S., Soares, P., Mantas, V., Göttsche, F.-M., & Trigo, I. (2020). Google Earth Engine open-source code for Land Surface Temperature estimation from the Landsat series. *Remote Sensing, 12*(9), 1471.

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

Liu, L., & Zhang, Y. (2011). Urban Heat Island Analysis Using the Landsat TM Data and ASTER Data: A Case Study in Hong Kong. *Remote Sensing, 3*(7), 1535-1552.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765-4774.

Mutani, G., & Todeschi, V. (2020). The effects of green roofs on outdoor thermal comfort, urban heat island mitigation and energy savings. *Atmosphere, 11*(2), 123.

Mutani, G., Cristino, V., & Bullita, M. (2018). The Urban Heat Island of the Metropolitan City of Turin: Strategies for a Sustainable Urban Planning. *Proceedings of the XXIII Convegno A.I.P.T. (Associazione Italiana Proprietà Termofisiche), Turin, Italy, 21–22 September 2017*, (pp. 105-123).

Mutani, G., Scalise, A., Sufa, X., & Grasso, S. (2024). Synergising Machine Learning and Remote Sensing for Urban Heat Island Dynamics: A Comprehensive Modelling Approach. *Atmosphere, 15*(12), 1435.

Mutani, G., Todeschi, V., Kämpf, J., Coors, V., & Fitzky, M. (2018). Building Energy Consumption Modeling at Urban Scale: Three Case Studies in Europe for Residential Buildings. *Proceedings of the 2018 IEEE International Telecommunications Energy Conference (INTELEC), Turin, Italy, 7–11 October 2018*, (pp. 1-8).

Oke, T. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society, 108*(455), 1-24.

Peng, S., Piao, S., Ciais, P., Friedlingstein, P., Ottle, C., Bréon, F.-M., . . . Myneni, R. (2012). Surface urban heat island across 419 global big cities. *Environmental Science & Technology, 46*(2), 696-703.

Slavcheva, R., Shabou, S., Mackres, E., & Bekkers, A. (2025, August 20). *Europe's Soaring Heat and the Great Air Conditioning Dilemma*. Retrieved August 28, 2026, from World Resources Institute: https://www.wri.org/insights/europes-heat-and-air-conditioning-dilemma

Stewart, I., & Oke, T. (2012). Local Climate Zones for urban temperature studies. *Bulletin of the American Meteorological Society, 93*(12), 1879-1900.

Voogt, J., & Oke, T. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment, 86*(3), 370-384.

Vujovic, S., Haddad, B., Karaky, H., Sebaibi, N., & Boutouil, M. (2021). Urban Heat Island: Causes, Consequences, and Mitigation Measures with Emphasis on Reflective and Permeable Pavements. *CivilEng, 2*(2), 459-484.

Wei, L., & Sobrino, J. (2024). Surface urban heat island analysis based on local climate zones using ECOSTRESS and Landsat data: A case study of Valencia city (Spain). *International Journal of Applied Earth Observation and Geoinformation, 130*, 103875.

Xu, H., Lin, D., & Tang, F. (2013). The Impact of Impervious Surface Development on Land Surface Temperature in a Subtropical City: Xiamen, China. *International Journal of Climatology, 33*(8), 1873-1883.

Zhang, J., Wang, Y., & Li, Y. (2006). A C++ program for retrieving land surface temperature from the data of Landsat TM/ETM+ band6. *Computers & Geosciences, 32*(10), 1796-1805.

Zhang, Y., Li, D., & Liu, L. (2022). Multi-year Landsat analysis of surface urban heat island dynamics. *Science of the Total Environment, 838*, 156456.

Zhou, Y., & Mutani, G. (2026). Downscaling land surface temperature and evaluating high-albedo pavements in Turin. *Sustainability, 18*(1), 1-24.

Appendix

Appendix A - Google Earth Engine

This appendix prsents the Google Earth Engine script used to generate the Land Surface Temperature, SUHII, and other predictor rasters described in Section 3.4. The seven processing stages are summarised in Tables 3.4 and 3.5. The same script produces the data for all four seasonal exports, with only the date window and if required the pickDate changed between each season run. It produces nine GeoTIFF layers- LST, SUHII, NDVI, NDBI, NDMI, NDWI, PVI, albedo and emissivity at 30m resolution in EPSG:32632. Available at the link: Google Earth Engine- LST Calculation.

Appendix B - Predictor Screening, Random Forest Modelling, and Mitigation Scenario Scripts

Each modelling script tunes its Random Forest models the same way: RandomizedSearchCV searches over the number of trees, maximum tree depth, minimum samples per leaf, and the number of features considered at each split, scored on five-fold cross-validated R^2. The best configuration from each search is then refit on the full training set to produce the final model. For the final fifteen-predictor models, the selected hyperparameters are reported in Table 5.5, and both the held-out test-set metrics (R^2, RMSE, MAE) and the five-fold cross-validated R^2 appear alongside them in Table 5.6, for every season and the All Seasons Combined configuration.

B.1 Predictor screening and full 22-predictor model –

Available at: ML_22_Predictors_Model.py

Runs Pearson correlation, predictor intercorrelation, sensitivity analysis, and VIF

screening across the initial 22-candidate predictor set, together with a tuned Random Forest fit and SHAP analysis on the full set (30-iteration search). Its VIF and correlation outputs underpin the predictor reduction described in Section 5.4.

B.2 Final 9-predictor seasonal models –

Available at: ML_9 Predictor Model

Repeats the tuning, evaluation, and SHAP pipeline on the reduced 15-predictor set (seven morphology, six climate, two spectral — NDBI and five others excluded for high VIF), executed in sequence for summer, winter, spring, and autumn. These are the models behind Chapter 5- 9 predictors.

B.3 All Seasons Combined model –

Available at: All_Seasons_Combined_Model

Combines the four seasonal datasets into one dataset and reapplies the same screening, tuning, and SHAP pipeline to it, producing the All Seasons Combined row of Tables in Chapter 5 and the figures reproduced in Appendix C.

B.4 SHAP based Mitigation modelling

Available at: SHAP_mitigation_feasibility

Appendix C – Additional Maps and Per- Season Statistics Tables

This appendix collects supplementary correlation, sensitivity and SHAP figures from the fifteen-predictor seasonal models and the All Seasons Combined

model that were referenced from Chapter 5 but not reproduced there in the interest of a readable main text. The seasonal Random Forest models, hyperparameter search and evaluation procedure are described in Sections 5.5 and 5.6; the feature-importance and SHAP interpretation are described in Section 5.7.

C.1 SHAP-Based Mitigation Analysis — Additional Seasons

This section provides the feature importance bar charts, SHAP beeswarm plots, and residual diagnostic figures for Winter, Spring, and Autumn 2025, following the same analytical pipeline as the Summer analysis presented in Chapter 6. Table D.3.1 below summarises the complete set of turning points and per-season mitigation potential for reference alongside the figures.

Winter 2025

In winter 2025, NDVI (mean |SHAP| = 0.492 °C) and Albedo (0.368 °C) jointly lead SUHI importance, with ISF in third place (0.191 °C). The winter turning points are: NDVI = 0.437, ISF = 0.607, Albedo = 0.126. Albedo emerges as the primary driver in several priority winter zones (Villaggio Snia 88 %, Tre Tetti — Bellezia 72 %, Vallette — Saffarona 65 %), where reflective surface measures are the indicated primary intervention. Four predictors show sign reversals between their marginal sensitivity and their Pearson correlation: sky view factor, population density, relative building height, and height-to-width ratio.

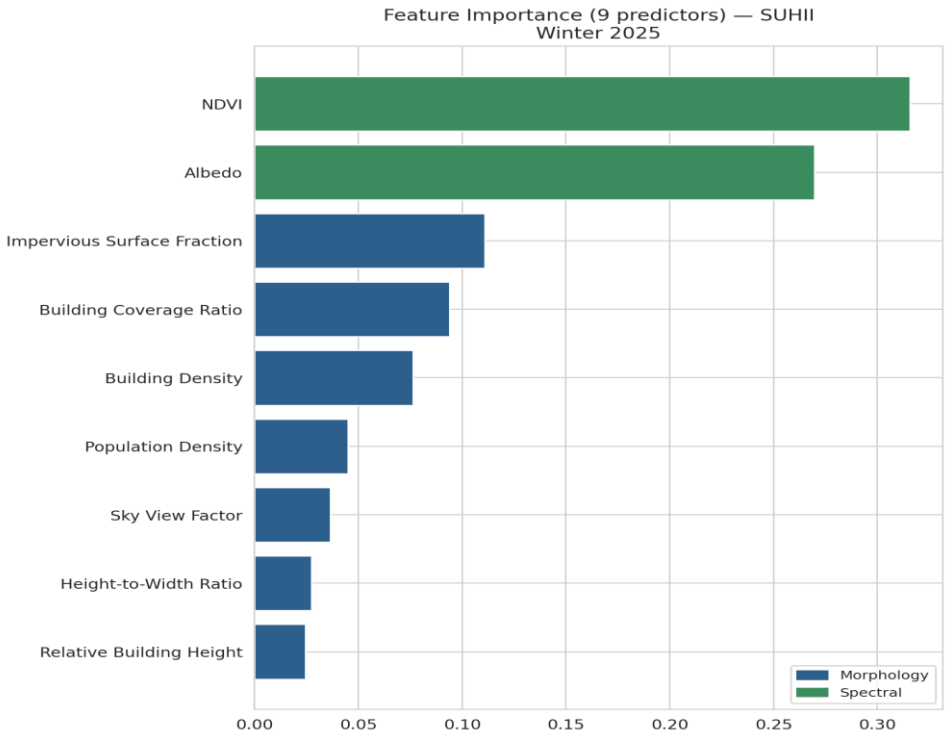


Figure C.15. Gini feature importance for Winter SUHII and SUHI; nine predictors.

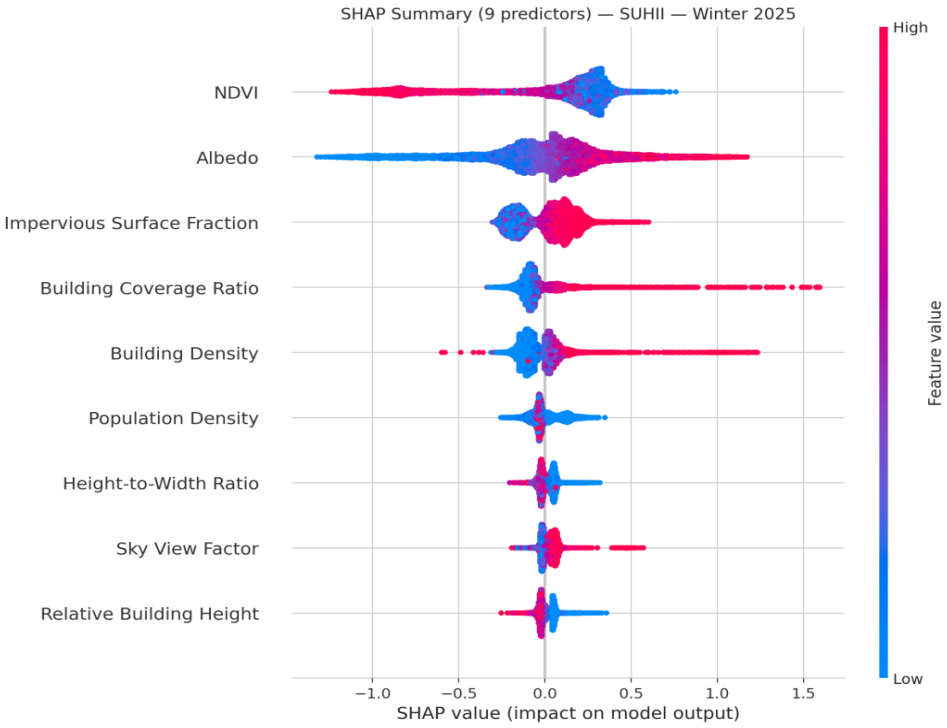


Figure C.16. SHAP beeswarm — Winter SUHII, showing per-cell predictor contributions.

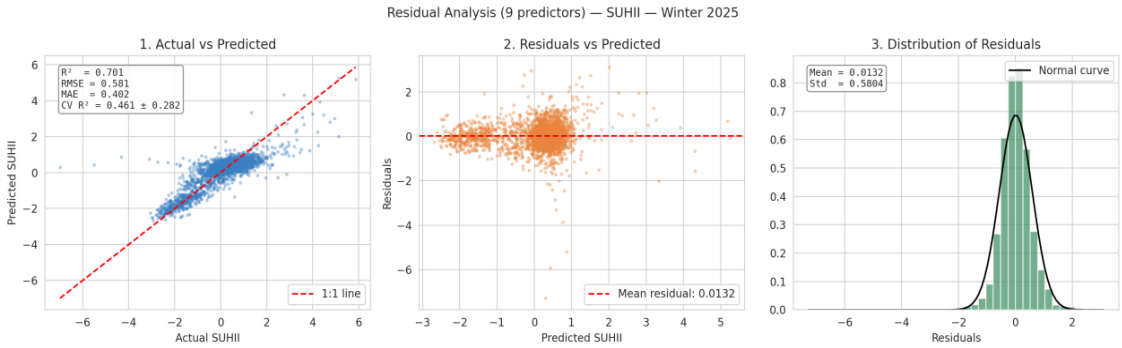


Figure C.17. Residual diagnostics — Winter SUHII: actual vs predicted, residuals vs predicted, and residual distribution.

Spring 2025

Spring 2025 shows NDVI (0.743 °C) and Albedo (0.526 °C) leading, with ISF third (0.402 °C). Spring turning points: NDVI = 0.405, ISF = 0.565, Albedo = 0.166. Spring carries the highest Most Emergency cell count (787), reflecting an intermediate vegetation state where even moderate NDVI deficits produce substantial SUHI contributions. Two predictors show sign reversals: height-to-width ratio and sky view factor.

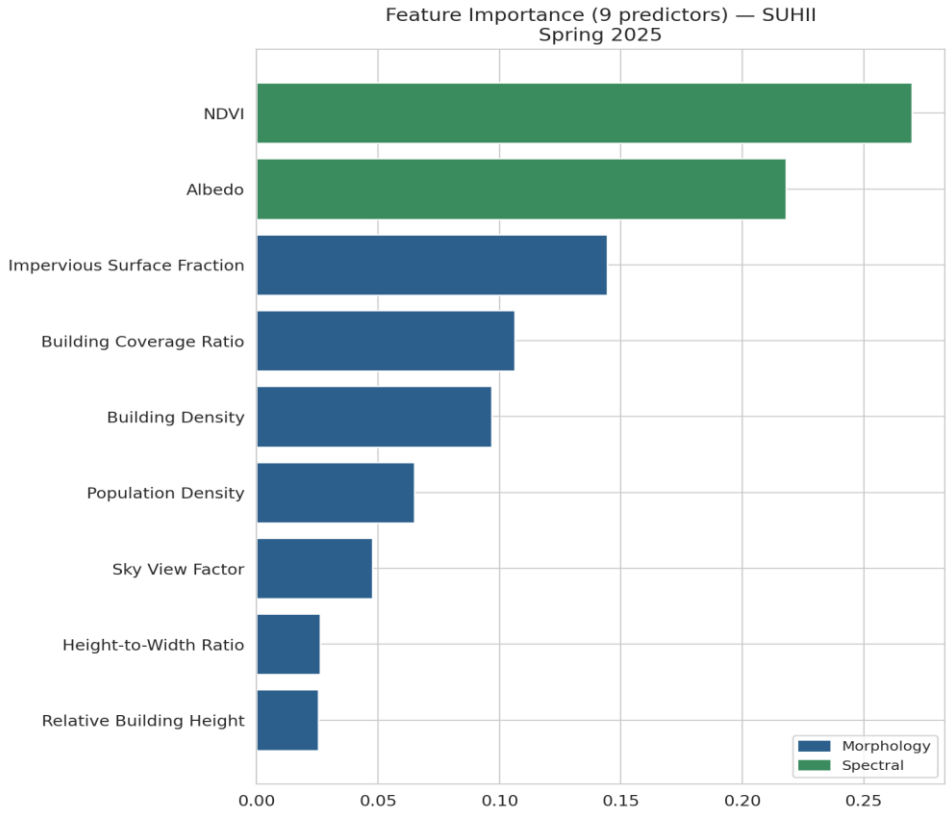


Figure C.18. Gini feature importance for Spring SUHII and SUHI; nine predictors.

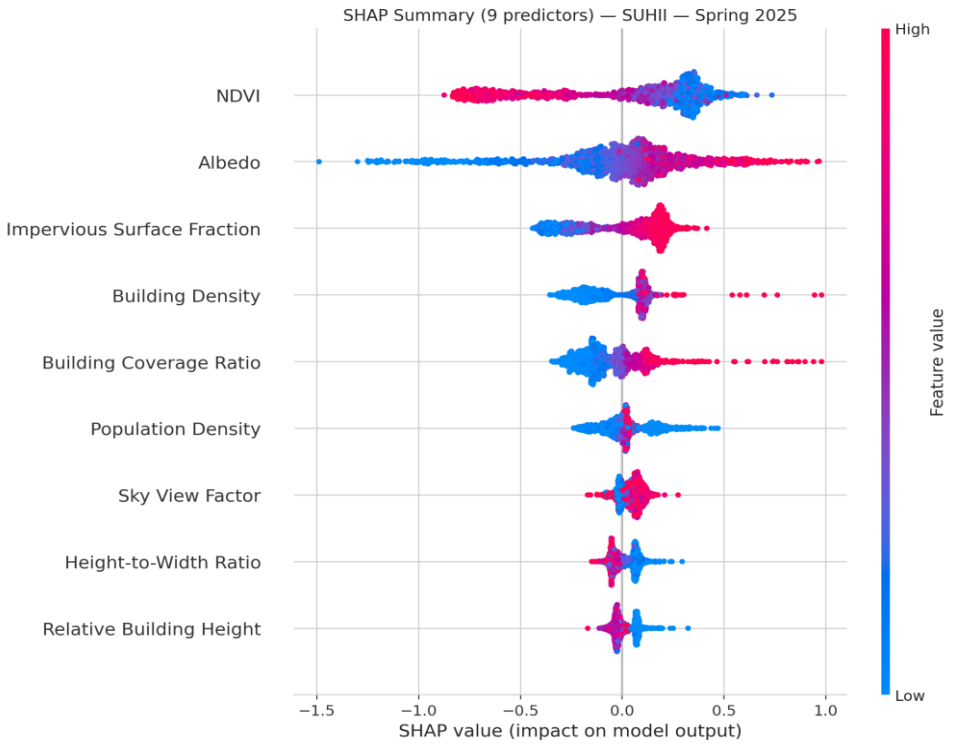


Figure C.19. SHAP beeswarm — Spring SUHII, showing per-cell predictor contributions.

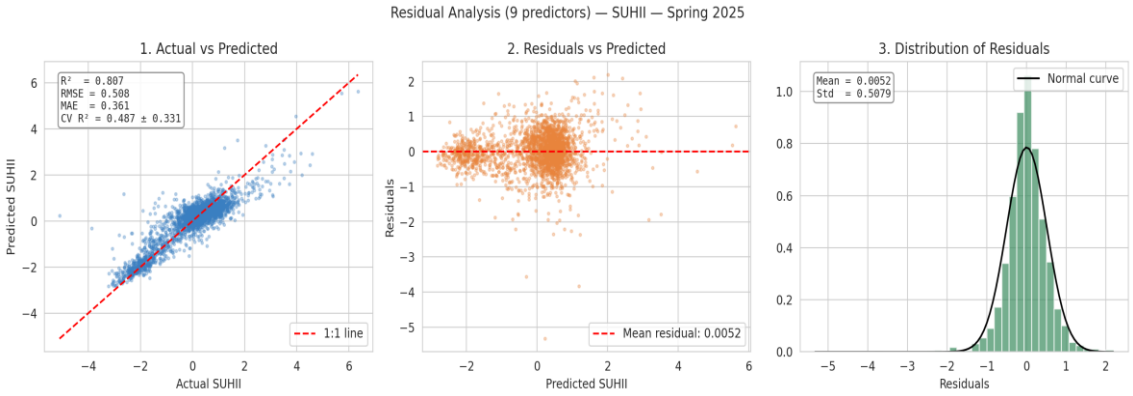


Figure C.20. Residual diagnostics — Spring SUHII: actual vs predicted, residuals vs predicted, and residual distribution.

Autumn 2025

Autumn 2025 shows NDVI leading (mean |SHAP| = 0.372 °C SUHII equivalent), with ISF second and Albedo third. Autumn turning points: NDVI = 0.574, ISF = 0.565, Albedo = 0.208. Autumn's NDVI turning point (0.574) is the highest of the four seasons, consistent with the dense canopy at peak senescence. Two predictors show sign reversals: height-to-width ratio and sky view factor, the lowest count across the four seasons.

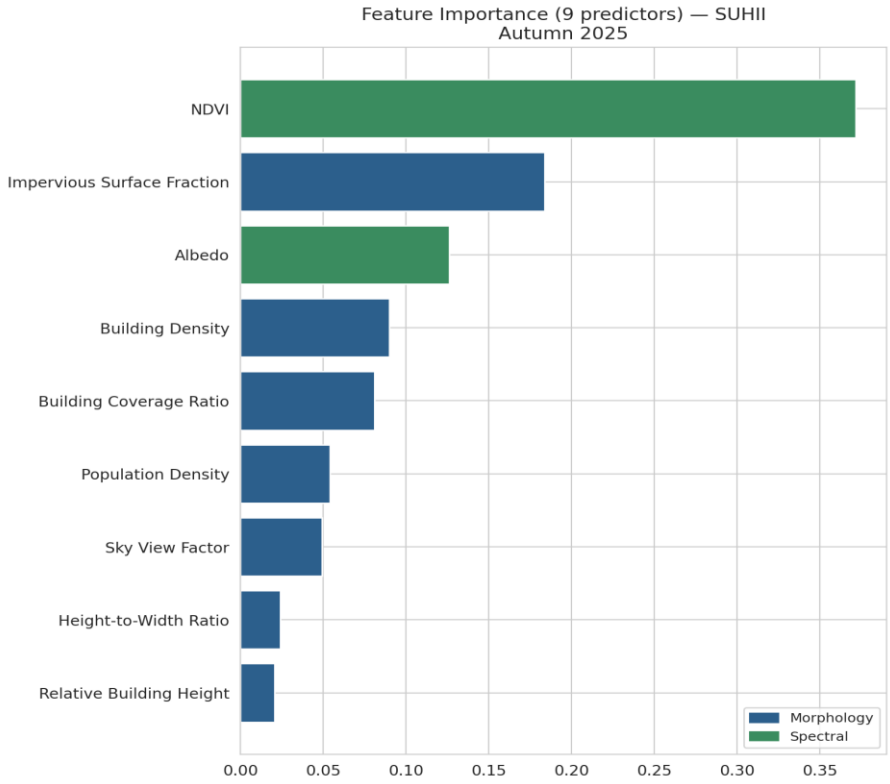


Figure C.21. Gini feature importance for Autumn SUHII and SUHI; nine predictors.

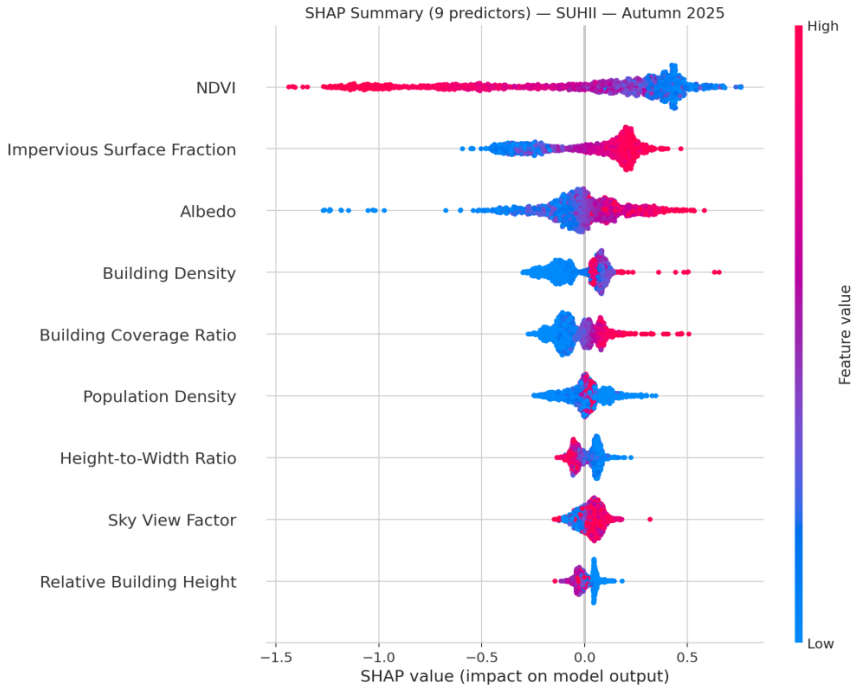


Figure C.22. SHAP beeswarm — Autumn SUHII, showing per-cell predictor contributions.

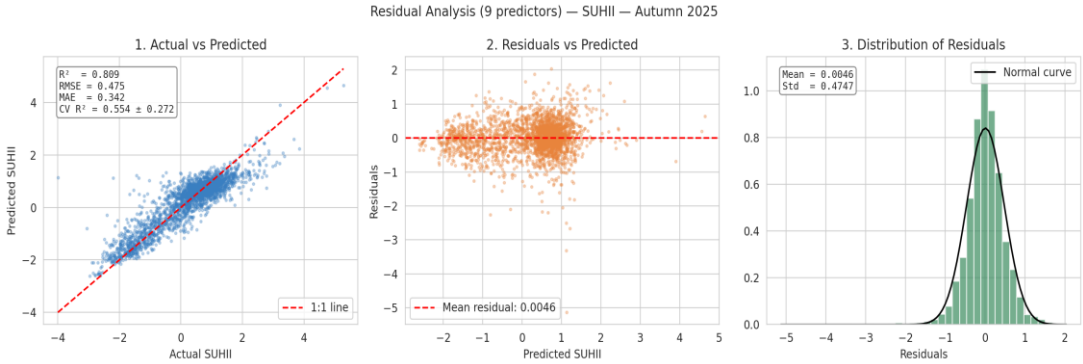


Figure C.23. Residual diagnostics — Autumn SUHII: actual vs predicted, residuals vs predicted, and residual distribution.