Sciences

Mapping and Modelling the Urban Heat Island in Northern Nairobi:

A Spatial Regression Analysis of Environmental and Morphological Drivers

Author: Beatrice Galetti (AUC), beatrice.galetti@student.auc.nl

Supervisor: Eric Koomen (VU, AUC), e.koomen@vu.nl

Reader: Eduardo Simao Da Graca Dias (VU, AUC), e.simaodagracadias@vu.nl

Tutor: Scott Dalby

26/05/2026

Major: Sciences

Word Count: 9033

Abstract

This study investigates the spatial variation and environmental drivers of nocturnal urban heat island (UHI) intensity across northern Nairobi, Kenya, using mobile air temperature transects and spatial regression modelling. Air temperature data were collected across four measurement sessions between August and October 2025, and UHI intensity was calculated as the difference between urban transect temperatures and a simultaneous rural reference at the Dagoretti meteorological station. Ordinary least squares (OLS) regression, spatial autocorrelation diagnostics, and spatial autoregressive (SAR) modelling were applied to examine the relationship between UHI intensity and six environmental and morphological predictors: NDVI, elevation, built-up surface, built-up height, population density, and local climate zones. Results show that mean nocturnal UHI intensity ranged from 1.64°C to 4.81°C across sessions, with peak values exceeding 6°C. NDVI emerged as the most consistent and significant predictor across all accepted sessions, with negative SAR coefficients ranging from $\beta = -0.960$ to $\beta = -2.098$, confirming the cooling role of vegetation cover. Elevation was a strong predictor in the OLS models, reflecting Nairobi's pronounced topographic gradient. Built-up variables did not retain significance in the final SAR models, likely reflecting both the difference between air temperature and land surface temperature as dependent variables and the limited coverage of the densest urban areas by the transect routes. SAR models consistently outperformed OLS, with pseudo-R^2 values between 0.698 and 0.892 compared to OLS R^2 values between 0.522 and 0.724, confirming the importance of accounting for spatial autocorrelation in urban temperature modelling. A continuous UHI intensity map was produced through spatial extrapolation of the September 5th OLS model. The findings contribute to the limited body of air temperature-based UHI research in African cities and highlight vegetation preservation and topographic context as key considerations for urban heat mitigation in Nairobi.

List of Abbreviations:

Abbreviation	Full Term
AIC	Akaike Information Criterion
ANBH	Average Net Building Height
CBD	Central Business District
DEM	Digital Elevation Model
GEE	Google Earth Engine
GHSL	Global Human Settlement Layer
GPS	Global Positioning System
KNN	K-Nearest Neighbours
LCZ	Local Climate Zone
LM	Lagrange Multiplier
LST	Land Surface Temperature
LULC	Land Use and Land Cover
NDVI	Normalised Difference Vegetation Index
OLS	Ordinary Least Squares
SAR	Spatial Autoregressive Model
SRTM	Shuttle Radar Topography Mission
UHI	Urban Heat Island
UTM	Universal Transverse Mercator
VIF	Variance Inflation Factor

Keywords: Urban heat island · Mobile transects · Spatial autoregressive model · NDVI · Nairobi · Nocturnal air temperature · East Africa · Elevation · Spatial autocorrelation · Urban vegetation

1. Introduction:

The Urban Heat Island (UHI) effect refers to the tendency for urban areas to experience higher temperatures than their rural surroundings due to the materials and structure of the built environment (Oke, 1982). The replacement of natural surfaces with impervious materials such as asphalt and concrete increases heat absorption and storage, while reduced vegetation limits cooling through evapotranspiration (Oke, 1982; Mohajerani et al., 2017). In addition, the geometry of urban areas, often described as the urban canyon effect, restricts airflow and traps outgoing radiation between buildings, further intensifying heat accumulation (Oke, 1981). As a result, urban areas experience elevated air temperatures, particularly during the evening when stored heat is released (Oke, 1981). This warming increases energy demand for cooling (Li et al., 2019), raises the risk of heat-related morbidity and mortality (Heaviside et al., 2017; Kassomenos & Begou, 2022), and reduces thermal comfort in everyday urban environments (Morris et al., 2017).

Despite significant growth in UHI research over the last two decades, three important gaps remain in the literature. First, the vast majority of studies focus on cities in Europe, North America, and Asia, leaving African cities severely underrepresented (de Almeida et al., 2021). This is particularly concerning given that population exposure to extreme heat is projected to increase more rapidly in East Africa than in any other world region (Liu et al., 2017). Second, the few studies conducted on African cities rely predominantly on satellite-derived land surface temperature (LST), which measures the radiative temperature of rooftops and road surfaces. This differs fundamentally from the near-surface air temperature experienced by people at ground level, which is the metric most directly linked to human health and thermal comfort (Sheng et al., 2017). Third, most studies investigating which urban factors drive UHI intensity rely on simple linear regression or correlation analyses (de Almeida et al., 2021), which assume spatial independence between observations, an assumption that geographic data rarely satisfies and that, when ignored, produces unreliable model estimates (Anselin, 1988).

This study aims to address all three gaps. The focus is Nairobi, Kenya, a rapidly growing equatorial city of over four million inhabitants (Mundia & Aniya, 2005; KNBS, 2025). It is characterized by stark contrasts between dense informal settlements, commercial districts, and vegetated suburbs, making it a compelling case study for UHI dynamics in sub-Saharan Africa (Simwanda et al., 2019). Air temperature data were collected using mobile transects, where a vehicle-mounted sensor recorded temperature at regular intervals along similar routes across the northern area of the city, capturing the spatial variation in near-surface air temperature that satellite imagery cannot provide. These measurements are used to map UHI intensity across northern Nairobi and to identify which environmental and morphological factors contribute most strongly to it. To do so, spatial regression models are applied that explicitly account for spatial autocorrelation. This study therefore addresses the following research question: How does urban heat island intensity vary spatially across northern Nairobi, and which environmental and morphological factors best explain this variation?

This overarching question is addressed through three sub-questions: (1) What is the spatial distribution and magnitude of near-surface air temperature differences between urban and rural areas in northern Nairobi, as measured by mobile transects? (2) To what extent do the environmental and morphological variables explain spatial variation in UHI intensity, when accounting for spatial autocorrelation? (3) Which of these factors are the strongest predictors of UHI intensity in the study area?

The remainder of this thesis is structured as follows. Section 2 reviews the existing literature on UHI dynamics, measurement approaches, and analytical methods, including the research gaps this study addresses. Section 3 describes the study area, data collection process, and analytical methodology. Section 4 presents the results, which are then discussed alongside concluding reflections, limitations, and directions for future research in Section 5.

2. Research Context

2.1 UHI Research and the African Context

The study of urban heat islands dates back to Luke Howard's early nineteenth century observations of elevated temperatures in London compared to its surrounding countryside (Howard, 1833). It was not until the work of Oke (1981, 1982) that the physical mechanisms behind the UHI were formally established, laying the foundation for the systematic research that followed. Over the past two decades, UHI research has grown substantially, driven by advances in remote sensing and increasing concern over urban heat as a public health issue (Voogt & Oke, 2003). However, this growth has been uneven: studies remain heavily concentrated in cities in Europe, North America, and East Asia, with Africa accounting for only a marginal share of the total literature (de Almeida et al., 2021).

This is particularly concerning given that Africa is undergoing some of the fastest urban growth in the world (United Nations, 2019). Much of this growth has been rapid and unplanned, producing cities dominated by dense informal settlements and limited green infrastructure (Simwanda et al., 2019). Furthermore, population exposure to extreme heat is projected to increase more rapidly in East Africa than in any other world region (Liu et al., 2017), making the lack of research in this area a significant concern. This underrepresentation likely reflects the limited availability of dense ground-based monitoring networks, long-term meteorological records, and research funding in many African cities.

Nairobi exemplifies this trajectory. As the capital of Kenya, it is one of East Africa's largest and fastest-growing cities, having undergone rapid and largely unplanned urban expansion over recent decades (Mundia & Aniya, 2005). Despite this, dedicated UHI research in Nairobi remains scarce, representing a gap this study directly addresses.

2.2 Measurement Methods in UHI Research

UHI intensity is most commonly measured by comparing temperatures in urban areas to those in a surrounding rural baseline (Voogt & Oke, 2003). Two types of temperature data are used for this: land surface temperature (LST) derived from satellites, and air temperature recorded either by fixed meteorological stations or through in-situ mobile measurements. LST captures the radiative temperature of rooftops and road surfaces, providing broad spatial coverage. However, it measures surface rather than air temperature, and Sheng et al. (2017) show that these two measures of UHI intensity are not directly comparable. Voogt and Oke (2003) further show that the difference between surface and air temperature can be substantial, particularly in cities where built-up heights vary considerably across the urban fabric. Air temperature is therefore a more direct indicator of the thermal conditions experienced by people in urban environments (Sheng et al., 2017). Fixed meteorological stations measure near-surface air temperature, but conventional station networks are typically designed to monitor regional atmospheric conditions rather than urban climate at the city scale, and their spacing is too coarse to capture the fine-scale thermal variation that characterises diverse urban environments (Muller et al., 2013). Mobile transects address this by mounting sensors on vehicles and recording air temperature continuously along predetermined routes, providing spatially detailed, human-scale measurements that neither satellites nor sparse station networks can produce (Romero Rodriguez et al., 2020). Despite these advantages, mobile transect-based UHI studies remain relatively uncommon in the literature compared to satellite and station-based approaches (de Almeida et al., 2021).

An important distinction in the literature is between daytime and nighttime UHI. Daytime UHI is largely driven by solar radiation and surface absorption, while nighttime UHI reflects the prolonged release of heat stored in built materials after sunset (Oke, 1982). Nighttime UHI is particularly significant from a human health perspective, as prolonged nocturnal heat stress reduces the body's ability to recover from daytime heat exposure (Cheval et al., 2024). Most of the limited existing studies on African cities, including those on Nairobi, use daytime LST measurements (Eshetie, 2024; Simwanda et al, 2019; Mwangi et al, 2018), meaning that nighttime thermal conditions in these cities remain poorly documented.

In Nairobi specifically, the existing literature is limited not only in quantity but also in analytical depth. Most studies focus on mapping the spatial distribution and intensity of LST rather than statistically modelling its drivers. The most analytically advanced study found in the peer-reviewed literature is limited to the Upper Hill neighbourhood: Mwangi et al. (2018) used simple linear regression to relate Land Surface Temperature (LST) to vegetation and built-up indices. However, this study does not utilize ambient air temperature measurements, scale its findings to cover the wider city, or account for spatial autocorrelation in a spatial regression framework. Scott et al. (2017) used fixed in-situ sensors to show that near-surface air temperatures varied considerably across a small number of Nairobi neighbourhoods, but did not attempt to model the drivers of this variation at the city scale. A review of the existing literature found no studies using mobile transects to measure air temperature across Nairobi, representing a methodological gap this study aims to address by providing a more spatially extensive picture of nighttime air temperature in northern Nairobi than has previously been documented in the literature.

2.3 Statistical Methods in UHI Research

Understanding which factors drive UHI intensity requires relating temperature measurements statistically to environmental and morphological variables. The most common approach in the literature is simple linear regression or Pearson correlation analysis (de Almeida et al., 2021). These methods are straightforward and widely used, but they assume that observations are spatially independent of one another. In reality, temperature at any given location tends to be similar to temperatures nearby, a property known as spatial autocorrelation, first formalised by Tobler (1970). When spatial autocorrelation is present and not accounted for, regression coefficients become unreliable and the significance of predictors tends to be overstated (Anselin, 1988). Yin et al. (2018) confirmed this directly in a UHI context: testing for spatial autocorrelation in LST data from Wuhan, China, they found significant spatial clustering, and showed that OLS consistently over or underestimated the influence of key predictors compared to a Spatial Error Model. Li et al. (2020), studying 419 global cities, similarly found that a Geographically Weighted Regression model substantially outperformed OLS in explaining spatial variation in UHI intensity, confirming that the relationship between urban form and heat is spatially non-stationary and cannot be reliably captured by global regression models.

A related methodological consideration is how predictor variables are measured at each observation point. Rather than relying on a single pixel value, focal statistics aggregate values across a defined neighbourhood radius around each point, capturing the local spatial context that more accurately reflects the environment influencing temperature at that location. Rafiee et al. (2016), studying nocturnal UHI in Amsterdam, explicitly tested different neighbourhood radii for aggregating tree volume and found that the relationship between tree volume and UHI intensity varied substantially depending on the scale of aggregation, demonstrating that the choice of neighbourhood radius is an important methodological decision that can meaningfully affect model results.

Most existing studies on East African cities, including Simwanda et al. (2019) on Nairobi, Lagos, Addis Ababa and Lusaka, and the Nairobi-specific study by Mwangi et al. (2018), use correlation-based approaches, simple linear regression, or gradient analysis rather than spatial regression models that account for autocorrelation. A further limitation of most UHI studies is that results are presented only at measurement locations rather than as a continuous surface across the study area. Spatial interpolation addresses this by using measured point values to estimate temperatures at unsampled locations, producing a continuous map of UHI intensity across the wider study area. Szymanowski and Kryza (2012) demonstrated that combining regression modelling with spatial interpolation of residuals produces continuous UHI maps from sparse point measurements, though this approach has rarely been applied in African urban contexts. This study combines spatial regression, focal statistics aggregation, and spatial interpolation to map and model nighttime UHI intensity across northern Nairobi.

3. Methodology

3.1 Study Area

Nairobi is the capital and largest city of Kenya, located in East Africa at an elevation of around 1,795 metres above sea level. With an estimated population of approximately 4.9 million in 2025, it is one of the fastest-growing cities in East Africa, having undergone rapid and largely unplanned urban expansion over recent decades (KNBS, 2025). The city has a Köppen-Geiger Cfb classification, characterised by a mild oceanic climate with limited seasonal temperature variation, two distinct rainy seasons peaking in April and November, and an annual mean temperature of approximately 17.7°C (Beck et al., 2018; Simwanda et al., 2019; University of Cape Town, 2017). Despite its equatorial location, Nairobi's high elevation moderates temperatures considerably compared to lowland East African cities, and the city descends along a notable altitudinal gradient from the western highlands toward the eastern plains (Simwanda et al., 2019).

The study focuses on northern Nairobi, an area characterised by a diverse mix of urban environments including commercial districts, medium and high density residential neighbourhoods, and areas of significant green space, including the Sigiria and Karura forests. The transect routes cover the upper Central Business District and the north-western periphery of the city, traversing a range of neighbourhood types at varying elevations. This diversity of urban form makes northern Nairobi a suitable area for examining how different physical environments influence local temperature patterns. The spatial extent of the study area and the four transect routes used in the analysis are illustrated in Figure 1.

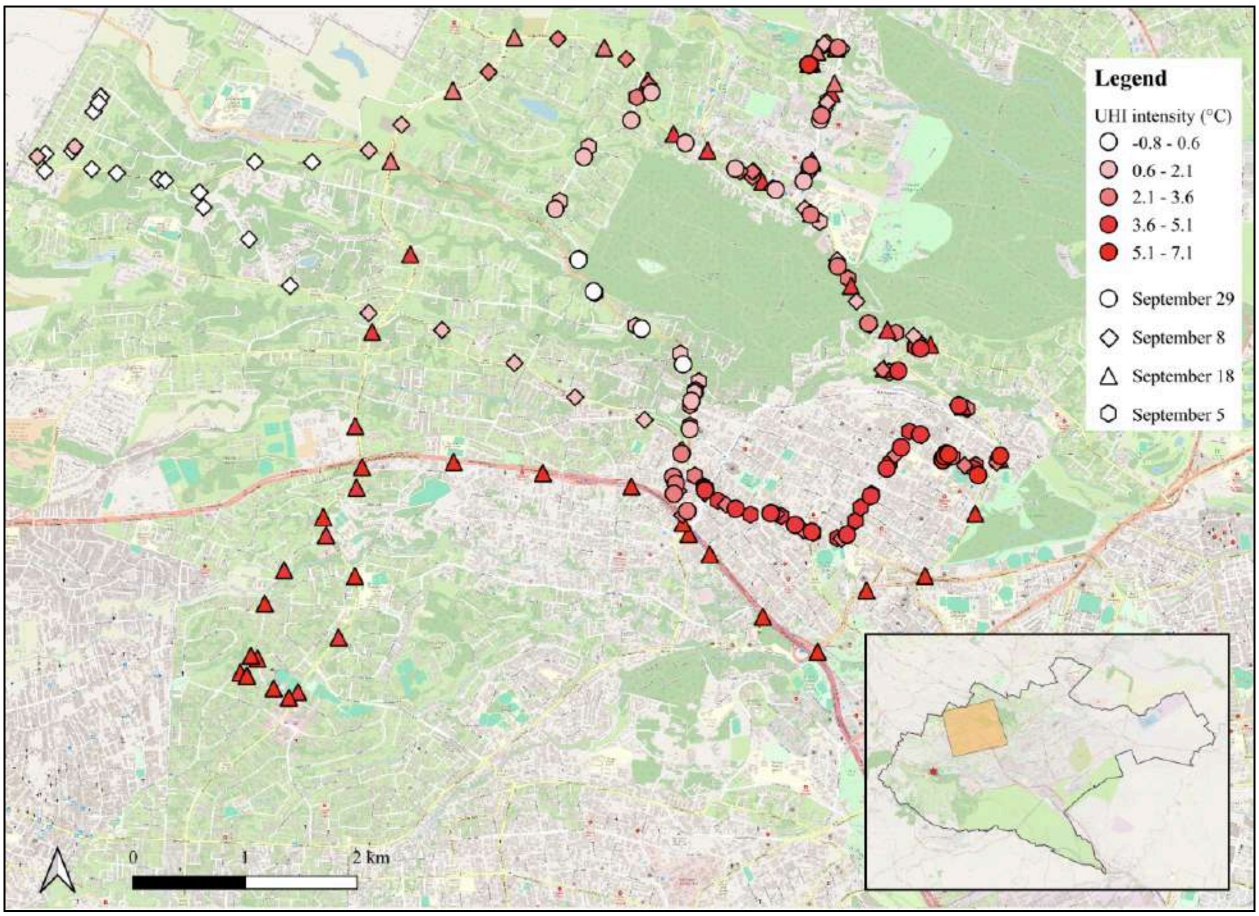

Figure 1: Spatial Extent of the Mobile Transect Routes and the Location of the Dagoretti Meteorological Station.

Note: The main map illustrates the study area containing the four mobile transect routes. The inset map shows the overall geographic outline of Nairobi, with a red marker indicating the reference location of the Dagoretti station and an orange shaded boundary defining the study area's placement within the wider city.

3.2 Data Collection

Temperature data were collected via mobile transects by Cordeiro (2025) across northern Nairobi during evening hours between 18:00 and 21:00. This time window was selected because nighttime UHI intensity is typically most pronounced during the hours following sunset, when heat stored in urban surfaces throughout the day is gradually released into the surrounding air (Oke, 1982). Measurements were conducted between late August and early October 2025, falling within Nairobi's dry season, a period of relatively low precipitation that reduces the likelihood of rainfall interfering with temperature recordings (University of Cape Town, 2017). A mobile temperature logger was mounted on the side view mirror of a vehicle, recording air temperature approximately every two minutes along predetermined circular routes covering major roads and various neighbourhoods across the study area. The vehicle was driven at an average speed of 35 to 40 km/h to prevent delays in temperature recording that can occur when driving too fast.

To isolate the urban contribution to local warming, a simultaneous rural reference temperature was recorded at the Dagoretti meteorological station, located on Nairobi's western periphery (Figure 1). Dagoretti provides a suitable rural baseline as it remains relatively less urbanised than the transect area, with higher vegetation cover and lower building density, and has been used as a reference in other urban temperature studies such as Scott et al. (2017). Since the station records temperature at hourly intervals, linear interpolation was applied to resample the reference data to one-minute intervals, aligning it with the transect measurements and avoiding abrupt transitions between hourly observations. UHI intensity was then calculated for each observation point using the standard urban-rural temperature difference formula (Voogt & Oke, 2003):

UHI = T_urban − T_ref

where T_urban is the air temperature recorded along the transect and T_ref is the simultaneous rural reference temperature at Dagoretti.

Of the nine sessions conducted, four were selected for spatial regression analysis: September 5th, 8th, 18th and 29th. These four sessions were the only ones lasting at least one hour, with the remaining sessions being shorter in duration and recording fewer observations, resulting in sparser spatial coverage of the study area. As shown in Table 1, the four selected sessions also capture a broad range of atmospheric conditions, from relatively dry and stable conditions on September 8th (relative humidity 51.4%, average wind speed 17.8 km/h) to more humid and windier conditions on September 18th (relative humidity 83.2%, average wind speed 24.1 km/h). Wind speed and humidity are known to influence UHI intensity, with higher wind speeds promoting heat dispersal and higher humidity affecting nocturnal heat release (Oke, 1982), and their variation across sessions will be reflected upon in the discussion. The

September 18th session included an unplanned interruption of approximately one hour, which may have affected the spatial consistency of temperature readings for that day and is acknowledged as a limitation.

Table 1: Summary of Descriptive Statistics and Atmospheric Parameters for Mobile Transect Observation Sessions

Date	Min UHI (°C)	Max UHI (°C)	Avg UHI (°C)	Min T (°C)	Max T (°C)	Avg T (°C)	Avg Wind Speed (km/h)	Avg Rel. Humidity (%)	N Points	Duration
20/08/2025	1.4	6.6	3.4	16.8	22.0	18.9	9.7	77.6	22	42 min
03/09/2025	1.9	5.6	3.5	20.2	23.5	21.7	15.9	62.2	18	34 min
05/09/2025	0.3	5.4	3.1	21.0	25.7	23.3	13.7	53.4	61	1 h
08/09/2025	-0.8	3.3	1.6	18.9	23.5	21.4	17.8	51.4	66	1 h 21 min
11/09/2025	1.6	3.8	2.8	20.2	22.7	21.5	15.7	63.5	15	14 min
16/09/2025	3.2	6.5	4.6	21.2	24.6	22.6	24.7	68.2	40	40 min
18/09/2025	3.0	6.4	4.8	19.1	23.3	21.5	24.1	83.2	61	2 h 29 min
29/09/2025	0.2	7.1	3.6	19.3	25.0	21.8	21.1	73.8	61	1 h
01/10/2025	2.1	6.2	4.1	21.5	26.2	24.1	16.1	51.3	44	43 min

Shaded rows indicate sessions selected for the regression analysis. T = near-surface air temperature.

3.3 Description of Explanatory Variables

To analyse the spatial variation of UHI intensity across northern Nairobi, a set of environmental and morphological variables was selected based on established UHI literature. These variables include NDVI, built-up surface, built-up height, population density, elevation, and local climate zones (LCZs). NDVI, built-up indicators, and elevation are among the variables most consistently associated with UHI intensity in the literature (de Almeida et al., 2021), while population density is commonly included as a proxy for urban compactness and anthropogenic heat emissions (Ramírez-Aguilar & Souza, 2019). LCZs, introduced by Stewart and Oke (2012) provide a standardized classification of urban form and thermal properties that has become increasingly common in UHI research, often as an alternative to the land use/land cover (LULC) variable (Johnson & Jozdani, 2019). Surface albedo is also frequently used in UHI studies for its relationship with the radiative properties of urban materials (de Almeida et al., 2021), but was excluded from all models in this study as it produced theoretically inconsistent associations with UHI intensity across sessions, alternating between positive and negative relationships, suggesting its effects could not be reliably isolated from confounding factors in this dataset. A complete overview of all variables, their sources, and spatial resolutions is provided in Table 2 in the Annex.

3.3.1 Environmental and Topographic Variables

Nairobi descends along a notable altitudinal gradient from the western highlands toward the eastern plains. Because air temperature naturally decreases with altitude as a result of the environmental lapse rate, elevation is a critical control variable in this study, distinguishing topographic cooling from urban-induced warming. Elevation data were obtained from the NASA Shuttle Radar Topography Mission (SRTM) Digital Elevation Model at 30m original resolution, accessed via Google Earth Engine (Gorelick et al., 2017), and resampled to 100m for the analysis. The spatial distribution of elevation across the study area is presented in Figure 2 (A).

Vegetation cover, represented through the NDVI, is another cooling factor. Vegetation mitigates UHI through shading, which reduces surface heating, and evapotranspiration, which converts sensible heat into latent heat and lowers ambient air temperature (Weng & Schubring, 2004). NDVI was derived from Sentinel-2 Harmonized Surface Reflectance imagery at 10m original resolution, accessed via Google Earth Engine (ESA; Gorelick et al., 2017). To produce a stable, representative layer, a median composite was generated from all September 2025 images with less than 20% cloud cover. NDVI was calculated using the standard normalised difference between the Near-Infrared (Band 8) and Red (Band 4) wavelengths (Tucker, 1979). The resulting NDVI layer is presented in Figure 2 (E).

3.3.2 Urban Morphology and Anthropogenic Variables

Built-up height and built-up surface area were obtained from the Global Human Settlement Layer (Pesaresi et al., 2024) at 100m resolution. Specifically, the Average Net Building Height (ANBH) layer from GHS-BUILT-H R2023 was used, which represents the average height of all buildings present within each 100m cell, excluding non-built structures such as vegetation or infrastructure. The built-up surface layer was taken from GHS-BUILT-S R2023, representing the total area covered by residential and non-residential rooftop structures combined within each cell, excluding roads, pavements, and other non-building impervious surfaces. The reference years are 2018 for built-up height (the most recent available) and 2025 for built-up surface. Taller structures contribute to the urban canyon effect by restricting nocturnal airflow and trapping long-wave radiation between building facades (Oke, 1981), while greater built-up surface area represents the conversion of natural land into thermal mass that absorbs solar energy during the day and releases it at night (Oke, 1981). The spatial distributions of both variables are presented in Figures 2 (C and F).

Population density was obtained from the GHS Population Grid R2023 (Pesaresi et al., 2024) at 100m resolution for the year 2025. It is included both as a proxy for urban compactness and as an indicator of anthropogenic heat emissions, capturing waste heat generated by concentrated energy consumption, traffic, and domestic activity (Ramírez-Aguilar & Souza, 2019; Manoli et al., 2019). The population density layer is presented in Figure 2 (D).

LCZs were integrated to capture neighbourhood-scale morphological characteristics that continuous variables alone cannot fully represent. The LCZ framework classifies urban and natural landscapes into 17 distinct classes based on building geometry, surface cover, and thermal properties, listed in Table 7 in the Annex (Stewart & Oke, 2012), enabling analysis of how specific neighbourhood types, such as compact low-rise or open high-rise, influence UHI intensity. The LCZ map for Nairobi was sourced from the WUDAPT community database, using the 100m resolution dataset generated by Moede (2021). The LCZ distribution across the study area is presented in Figure 2 (B).

All variables were reprojected to UTM Zone 37N (EPSG:32637) and resampled to a uniform 100m resolution in QGIS (version 3.38.2).

3.4 Neighbourhood Effect and Focal Statistics

The air temperature recorded at any given location is not solely determined by surface conditions at that exact point. Rather, it reflects the thermal characteristics of the surrounding landscape, which Oke (2004) and Stewart and Oke (2012) refer to as the source area: the spatial extent of surface that influences a temperature measurement at screen height. This is particularly relevant for mobile transect data, which captures near-surface air temperature at approximately 1.5m above ground level as the vehicle moves through the urban environment. At this height, the air mass integrates heat from surrounding surfaces rather than responding solely to the material directly beneath the sensor.

To account for the source area effect, focal statistics were applied to all continuous explanatory variables, including built-up height, built-up surface, population density, NDVI, and elevation, replacing individual pixel values with spatially smoothed estimates representing the thermal environment around each observation point. A circular neighbourhood of 150m radius (3x3 pixels at 100m resolution) was applied uniformly across all variables. This radius was selected as an intermediate value between the scale at which vegetation cooling most strongly influences air temperature, found to be 40 to 60m by Rafiee et al.

(2016), and the 200m scale at which built-up surface conditions most strongly influence nocturnal air temperature in dense urban environments (Nichol and Wong, 2008). Comparable mobile transect studies have applied radii of 250m following the source area recommendations of Stewart and Oke (2012) for characterising the surface environment around screen-height temperature observations (Gunn, 2023). All calculations were performed in UTM Zone 37N (EPSG:32637) to ensure distances were defined in metres.

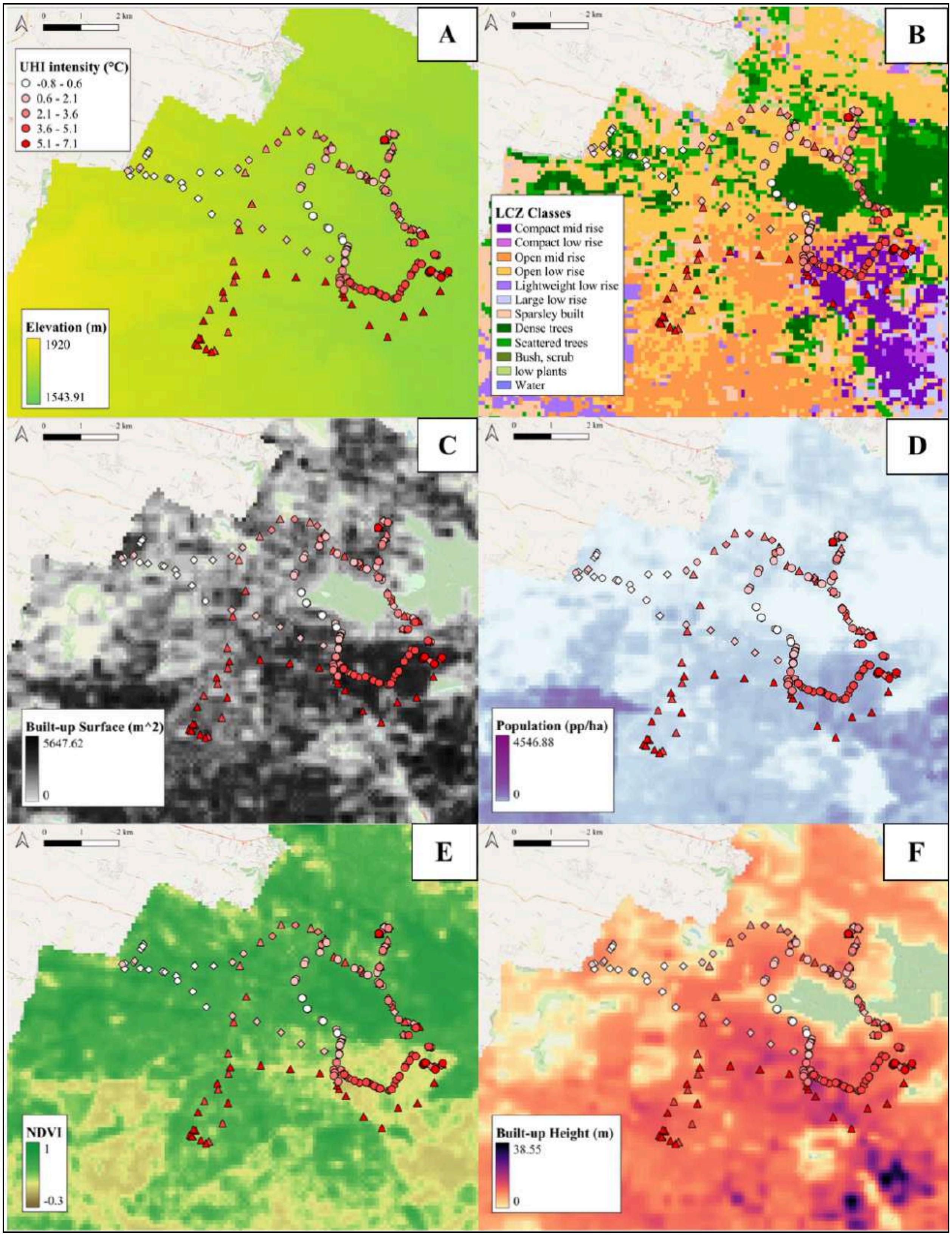

Figure 2: Spatial Distribution of the Six Explanatory Variables and Selected Mobile Transect Routes

3.5 Statistical Framework and Model Selection

Following variable preparation, a statistical analysis was applied to examine the relationship between urban form and observed UHI intensity across the four sessions selected for analysis: September 5th, 8th, 18th, and 29th. The analysis followed a sequential framework of OLS regression, spatial autocorrelation diagnostics, and spatial lag modelling, consistent with the approach applied in comparable UHI regression studies (Yin et al., 2018; Zeynali et al., 2026; Rafiee et al., 2016). All statistical analyses were conducted in R (version 4.3.1)

3.5.1 Data Integration and Pre-processing

Focal mean values were extracted at each GPS observation point, synchronising the explanatory variables with the temperature measurements. A frequency check was performed on the LCZ categories, and classes with fewer than five observations were excluded to prevent model overfitting on rare urban types, reducing the total number of observations available for analysis in each session. For September 5th, LCZ classes 8, 9, 11, and 14 were removed (7 observations total), leaving 54 observations for modelling. For September 8th, LCZ classes 3, 11, and 12 were removed (9 observations total), leaving 57 observations. For September 18th, LCZ classes 2, 8, 11, and 14 were removed (7 observations total), leaving 54 observations. For September 29th, LCZ classes 3, 8, 9, 11, and 12 were removed (11 observations total), leaving 50 observations. LCZ 6 (Open Low-rise) was retained as the reference category for dummy coding across all sessions, as it represented the most prevalent urban morphology in the sample.

3.5.2 OLS Regression and Multicollinearity Diagnostics

An initial OLS regression was run for each session to assess multicollinearity using the Variance Inflation Factor (VIF). When two or more variables exceed a VIF threshold of 10, their shared variance renders coefficient estimates unstable and unreliable (Anselin, 1988). In the September 8th and 18th sessions, all variables fell within acceptable limits. In the September 5th session, built-up height (VIF = 18.9) and built-up surface (VIF = 15.8) exceeded this threshold and were removed. In the September 29th session, built-up surface (VIF = 11.5) was removed prior to further analysis.

3.5.3 Spatial Autocorrelation Diagnostics

A fundamental assumption of OLS regression is that residuals are spatially independent. Violation of this assumption can produce unreliable coefficient estimates and overstated significance levels (Anselin, 1988), a problem frequently documented in urban temperature regression studies (Yin et al., 2018; Zeynali et al., 2026; Eshetie, 2024). To test this assumption, a spatial weights matrix was constructed using the K-Nearest Neighbours (KNN) algorithm, which represents a standard approach for modelling spatial dependence in geolocated point data (Kubara & Kopczewska, 2023; Anselin, 1988). As Kubara and Kopczewska (2023) note, there is no universally agreed value for k in the spatial econometric literature, and its selection is often somewhat arbitrary. A value of $k = 4$ was selected as a conservative specification to ensure that connectivity remained local given the relatively small sample sizes of 50 to 57 observations per session, thereby avoiding spatial lags becoming overly global and diluting local spatial structure.

With the spatial weights structure defined, a Global Moran's I test was performed on the residuals of the OLS models. This test evaluates whether residuals are randomly distributed or instead exhibit a geographic pattern. A significant positive Moran's I statistic indicates that the model systematically over- or under-estimates temperatures within specific spatial clusters, suggesting that standard linear regression is insufficient to capture the spatial structure underlying UHI intensity (Moran, 1950). Significant spatial autocorrelation was detected across all four sessions.

To determine the most appropriate alternative modelling approach, Lagrange Multiplier (LM) diagnostics were then applied. These tests compare two alternative forms of spatial dependence: the Spatial Lag Model, in which temperatures at a given location are directly influenced by neighbouring temperatures through spatial spillover effects, and the Spatial Error Model, in which spatial dependence exists only within the error term (Anselin, 1988). Across all sessions, the Spatial Lag statistic was more significant than the Spatial Error statistic, indicating that a Spatial Autoregressive (SAR) specification represented the more appropriate modelling framework.

3.5.4 Spatial Lag Model and Backwards Elimination

Based on the LM diagnostic results, the SAR model was implemented for each session. This model introduces a spatial lag parameter (rho) that captures the influence of neighbouring UHI values on each observation, thereby accounting for spatial dependence in the data (Anselin, 1988). Previous UHI studies have shown that SAR models can outperform conventional OLS regression when spatial autocorrelation is present, producing improved model fit and more robust coefficient estimates (Yin et al., 2018; Zeynali et al., 2026). The SAR model takes the form:

$$\mathbf{y = \rho Wy + X\beta + \varepsilon}$$

where y is the vector of UHI intensity values, ρ is the spatial autoregressive parameter, W is the spatial weights matrix, X is the matrix of predictor variables, β is the vector of regression coefficients, and ε is the error term.

Following SAR estimation, backwards elimination was conducted for each session, systematically removing the least significant variable at each step and re-running the model until all remaining variables were statistically significant at $p < 0.05$. The Akaike Information Criterion (AIC) was monitored at each step to confirm that model fit improved or was maintained with each variable removal (Akaike, 1974). Residual spatial autocorrelation was re-tested after each elimination step to confirm that spatial dependence remained adequately resolved throughout the process.

The September 29th session was excluded at this stage. Despite resolving multicollinearity, both the OLS and the SAR models for this day produced a significant positive NDVI coefficient, contradicting the well-established cooling effect of vegetation on near-surface air temperature. This logical inconsistency, rather than a purely statistical criterion, was the basis for exclusion. The three remaining sessions, September 5th, 8th, and 18th, were retained for the final SAR models and subsequent analysis.

The final SAR models for each session, alongside the corresponding OLS models for comparison, are presented in Table 3 in the results section. This side-by-side comparison illustrates how accounting for spatial autocorrelation alters coefficient estimates relative to OLS.

3.6 Spatial Extrapolation and Mapping

The final stage involved generating a continuous UHI intensity surface across the study area. Although the SAR model provides more accurate coefficient estimates by accounting for spatial autocorrelation, it cannot be directly applied to unsampled raster pixels (Anselin, 1988). The spatial lag parameter requires knowledge of neighbouring UHI values at unsampled locations, values that do not exist beyond the transect route. OLS regression coefficients were therefore used for the extrapolation, as they can be applied directly to continuous raster predictor layers without requiring neighbouring observations, following the approach applied in comparable air temperature mapping studies (Ninyerola et al., 2007; Rafiee et al., 2016).

The regression equation applied takes the form:

$$\mathbf{UHI} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \times \mathbf{X}_1 + \boldsymbol{\beta}_2 \times \mathbf{X}_2$$

where β_0 is the intercept and β_1 and β_2 are the unstandardized OLS coefficients for the significant variables (X_1 and X_2). To ensure predictions remained within the geographical scope of the empirical data, the extrapolation was spatially restricted to a 3km buffer around the transect routes. This distance was selected because it approximates the spatial extent within which the predictor variables remain broadly within the ranges observed along the transect. The equation was applied pixel-by-pixel to the 100m resolution focal mean raster layers in QGIS.

4. Results

4.1 Descriptive Statistics

Mean UHI intensity varied considerably across the four sessions, as shown in Table 1. The descriptive statistics for all variables across all four sessions are presented in Tables 4 and 5. September 18th recorded the highest mean intensity (4.81°C, max 6.36°C), while September 8th recorded the lowest (1.64°C). September 5th and 29th showed intermediate mean intensities of 3.21°C and 3.83°C respectively. September 8th was also the only session showing (slightly) negative values for a few observation points, indicating that the detected temperature in certain areas was lower than the rural reference.

As shown in Table 6, the sampled observations were dominated by LCZ 6 (Open Low-rise) across all sessions, with LCZ 5 (Open Mid-rise) and LCZ 2 (Compact Mid-rise) also consistently present. Elevation ranged from approximately 1670 to 1740m on average, and NDVI from 0.10 to 0.64, capturing both densely vegetated areas and heavily built areas. Mean built-up surface ranged from approximately 1971 to 2943 m² per 100m cell across sessions, mean built-up height from 9.31 to 10.85m, and mean population density from 33.54 to 75.02 persons per hectare, reflecting the varied urban densities encountered along the different transect routes. The descriptive statistics for all variables across all four sessions are presented in Tables 4 and 5.

4.1.2 Pairwise Correlations

All Pairwise Correlation Matrices are shown in Figure 3-6. NDVI showed a consistent negative relationship with UHI intensity across all four sessions ($r = -0.584, -0.620, -0.649, -0.256$ for September 5th, 8th, 18th, and 29th respectively), confirming that vegetated areas were cooler in every session. Elevation showed a strong negative relationship on September 5th, 8th, and 29th ($r = -0.539, -0.765, -0.745$), consistent with cooling at higher altitudes. However, September 18th showed a positive elevation correlation ($r = 0.147$), which contradicts the expected lapse rate effect and makes the variable unreliable. Population density and built-up variables showed consistently positive associations across sessions, particularly in September 18th where built-up height showed the strongest positive correlation ($r = 0.608$).

4.2 OLS Regression Results

The initial OLS models revealed significant relationships between UHI intensity and the predictor variables, though with notable differences across sessions. September 5th produced the strongest OLS fit ($R^2 = 0.724$), with both elevation ($\beta = -0.044$, $p < 0.001$) and NDVI ($\beta = -2.906$, $p = 0.023$) showing significant negative coefficients, consistent with expected cooling effects. September 8th also showed a strong elevation effect ($\beta = -0.018$, $p < 0.001$), though NDVI did not reach significance ($p = 0.418$). September 18th showed a weaker overall fit ($R^2 = 0.522$), with NDVI significant and negative ($\beta = -3.845$, $p = 0.006$) but elevation showing a positive coefficient ($\beta = 0.001$, $p = 0.726$). Since this positive elevation coefficient was not significant, it was not considered sufficient grounds for discarding the session, and was removed during backwards elimination without affecting the model. September 29th, however, produced a significant positive NDVI coefficient ($\beta = 5.38$, $p = 0.034$) across two independent model specifications, directly contradicting the well-established cooling effect of vegetation. This logical inconsistency, rather than a purely statistical criterion, was the basis for excluding September 29th from all further analysis.

4.3 Spatial Regression Results

Global Moran's I tests confirmed significant positive spatial autocorrelation in OLS residuals across all four sessions ($p < 0.001$ in all cases), and LM tests consistently favoured the Spatial Lag specification over the Spatial Error model, supporting the selection of the SAR model for all sessions.

A comparison of the final OLS and SAR model coefficients for the three accepted sessions is presented in Table 3. The SAR models were finalised through backwards elimination, retaining only variables that maximised the number of significant predictors and minimised AIC, while ensuring residual spatial autocorrelation remained resolved throughout (residual LM $p > 0.05$ in all final models). The spatial lag parameter was strongly significant in all three sessions ($\rho = 0.712$, 0.781, and 0.675 for September 5th, 8th, and 18th respectively, all $p < 0.001$), indicating that between 68 and 78% of the spatial variation in UHI intensity at any given location is associated with temperatures in the surrounding area rather than local surface characteristics alone.

Interestingly, NDVI was the only variable to remain significant across all three sessions in the final SAR models, with negative coefficients in every case ($\beta = -1.528$ for September 5th, $\beta = -0.960$ for September 8th, $\beta = -2.098$ for September 18th). This consistency across sessions with different routes and atmospheric conditions identifies vegetation cover as the most stable predictor of UHI intensity in northern Nairobi. Elevation retained a negative coefficient in both the September 5th and 8th final SAR models, although it did not reach significance at $p < 0.05$ in either session. LCZ 12 (Scattered Trees) also reached significance in the September 5th SAR model ($\beta = -0.521$, $p = 0.003$), suggesting that beyond vegetation density captured by NDVI, the spatial arrangement of tree cover contributes independently to localised cooling. However LCZ categories as a whole showed inconsistent coefficient signs across sessions, with some classes alternating between positive and negative associations on different days, limiting their overall interpretive value.

4.4 Spatial Extrapolation

The September 5th OLS model was selected as the basis for spatial extrapolation as it was the only session in which both elevation and NDVI produced statistically significant and physically consistent negative coefficients, and it achieved the highest OLS R^2 of the three accepted sessions (0.724). Applied across the masked study area, the extrapolation produced UHI values ranging from -1.27°C to 6.78°C, with a mean of 2.49°C and a standard deviation of 1.58°C. The final map can be visualized in Figure 7. The spatial pattern shows the highest UHI values concentrated in the southern and eastern zones characterised by lower vegetation cover and higher built-up density, while the vegetated residential areas and forest patches in the north and west remain consistently cooler. Given that the OLS model does not account for spatial spillover effects captured by the SAR parameter, the absolute values should be treated as indicative, with the spatial pattern being the more reliable output.

Table 3: Comparison of Unstandardized OLS and SAR Model Coefficients by Session, along with Other Model Statistics.

OLS and SAR Model Coefficients by Session
Dependent variable: UHI Intensity

	September 5		September 8		September 18	
Variable	**OLS**	**SAR**	**OLS**	**SAR**	**OLS**	**SAR**
Intercept	78.260***	16.988	33.496***	6.093*	3.024	2.409***
Elevation	-0.044***	-0.009	-0.018***	-0.003	0.001	—
NDVI	-2.906*	-1.528*	-1.317	-0.960*	-3.845**	-2.098**
Built Surface	—	—	0.000	—	0.000	—
Built Heights	—	—	0.001	—	0.084	—
Population	0.003	—	0.001	—	-0.003	—
LCZ 2 (Compact Mid-rise)	-0.386	0.046	-0.029	—	—	—
LCZ 5 (Open Mid-rise)	0.478	0.061	0.044	—	-0.012	—
LCZ 9 (Sparsely Built)	—	—	0.296	—	-0.120	—
LCZ 12 (Scattered Trees)	-0.750*	-0.521**	—	—	0.219	—
Rho (ρ)	—	*0.712****	—	*0.781****	—	*0.675****
Pseudo-R^2	—	*0.875*	—	*0.892*	—	*0.698*
R^2	*0.724*	—	*0.690*	—	*0.522*	—
Residual LM p-value	—	*0.057*	—	*0.279*	—	*0.546*
N	*54*	*54*	*57*	*57*	*54*	*54*

*Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. — indicates variable removed during backwards elimination, dropped due to multicollinearity (VIF > 10), or not present in sample for that session.*

Table 4: Descriptive Statistics of Spatial Variables for September 5 and September 8

	September 5				September 8			
Variable	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
UHI Intensity (°C)	3.21	1.09	0.68	4.98	1.64	1.14	-0.84	3.31
Elevation (m)	1696	13.62	1670	1738	1732	39.54	1676	1807
NDVI	0.34	0.13	0.10	0.57	0.41	0.15	0.11	0.64
Built-up Surface (m^2)	2570.9	1115.3	308.6	4528.4	2388.7	1150.9	60.6	4467.6
Built-up Heights (m)	10.01	3.06	3.05	15.92	9.31	4.04	1.28	21.02
Population (ppl/ha)	50.12	50.79	2.90	179.52	42.05	44.17	0.27	174.60

Table 5: Descriptive Statistics of Spatial Variables for September 18 and September 29

	September 18				September 29			
Variable	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
UHI Intensity (°C)	4.81	0.85	3.11	6.36	3.83	1.90	0.35	7.07
Elevation (m)	1740	38.55	1671	1790	1698	14.68	1676	1735
NDVI	0.40	0.11	0.12	0.57	0.35	0.16	0.10	0.64
Built-up Surface (m^2)	1970.8	782.9	470.8	3745.6	2943.4	1130.0	60.6	4467.6
Built-up Heights (m)	10.05	3.39	0.50	19.46	10.85	2.97	1.28	21.02
Population (ppl/ha)	33.54	25.80	4.24	113.64	75.02	58.27	0.27	171.88

Table 6: LCZ Classes Frequencies Table for All Sessions

	September 5			September 8			September 18			September 29		
LCZ Class	**N**	**%**	**Mean UHI**	**N**	**%**	**Mean UHI**	**N**	**%**	**Mean UHI**	**N**	**%**	**Mean UHI**
2 (Compact Mid-rise)	6	11.1	4.01	7	12.3	2.90	—	—	—	9	18.0	5.12
5 (Open Mid-rise)	11	20.4	4.03	6	10.5	2.60	18	33.3	5.35	19	38.0	4.18
6 (Open Low-rise)	31	57.4	3.02	39	68.4	1.28	21	38.9	4.44	22	44.0	2.99
9 (Sparsely Built)	—	—	—	5	8.8	1.49	10	18.5	4.71	—	—	—
12 (Scattered Trees)	6	11.1	1.91	—	—	—	5	9.3	4.65	—	—	—

All other LCZ classes were not present in the mobile transect routes.

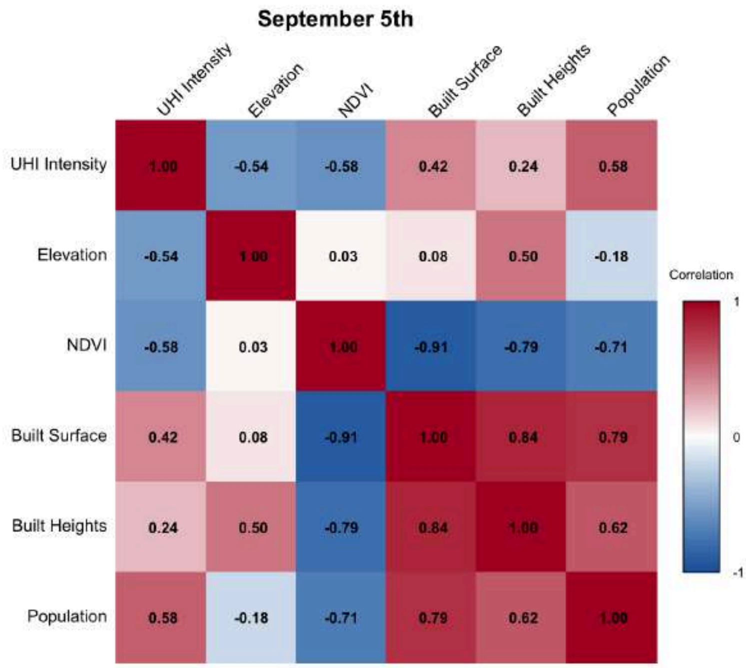

Figure 3: Pairwise Correlation Matrix for September 5th

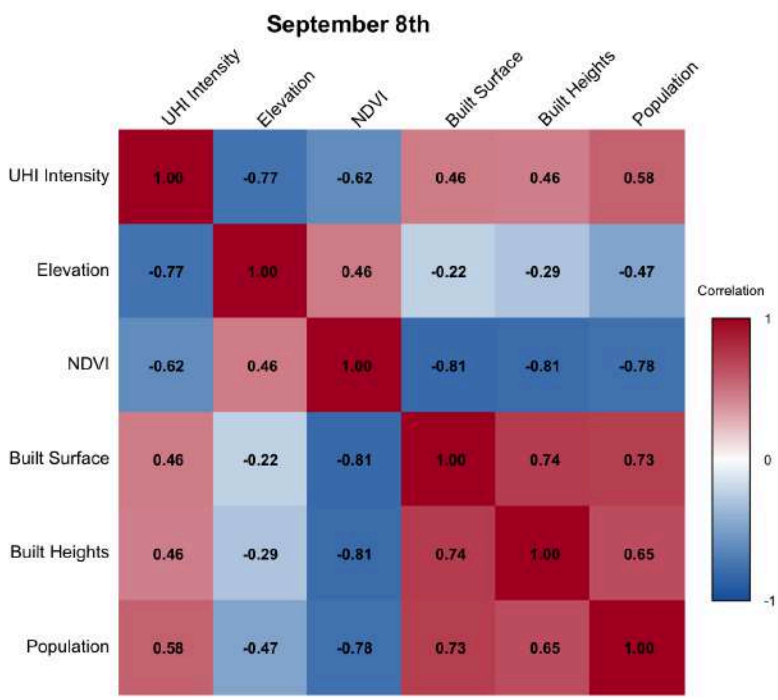

Figure 4: Pairwise Correlation Matrix for September 8th

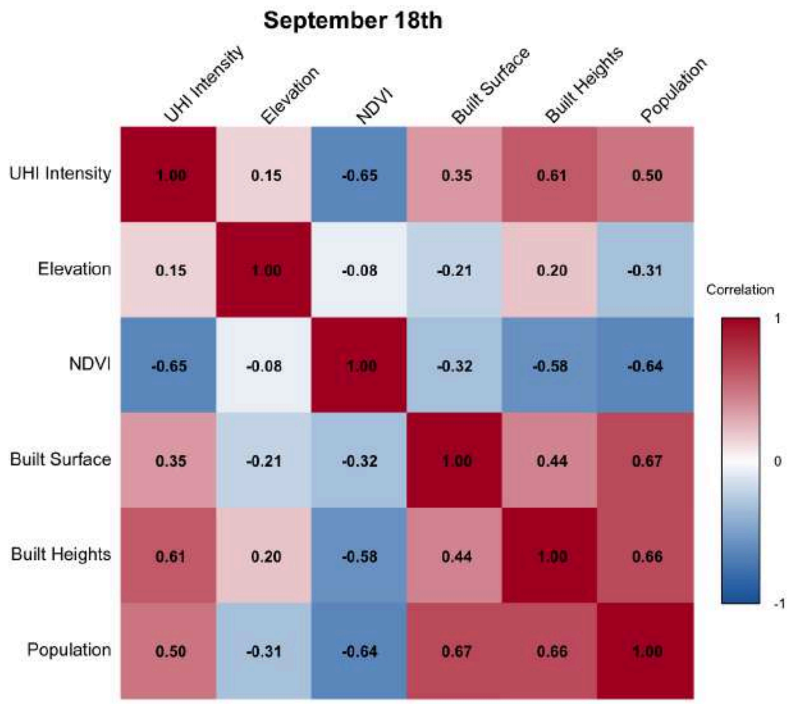

Figure 5: Pairwise Correlation Matrix for September 18th

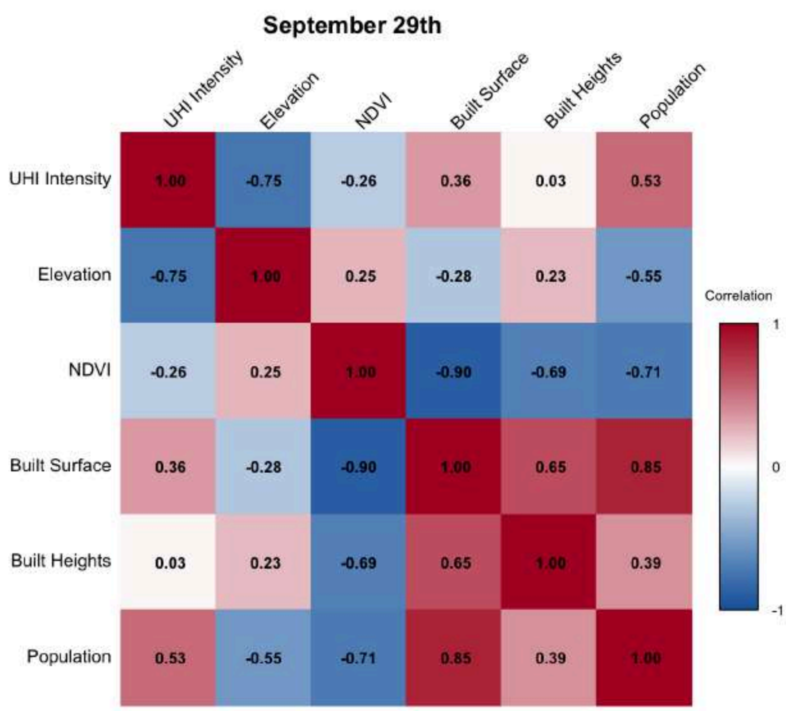

Figure 6: Pairwise Correlation Matrix for September 29th

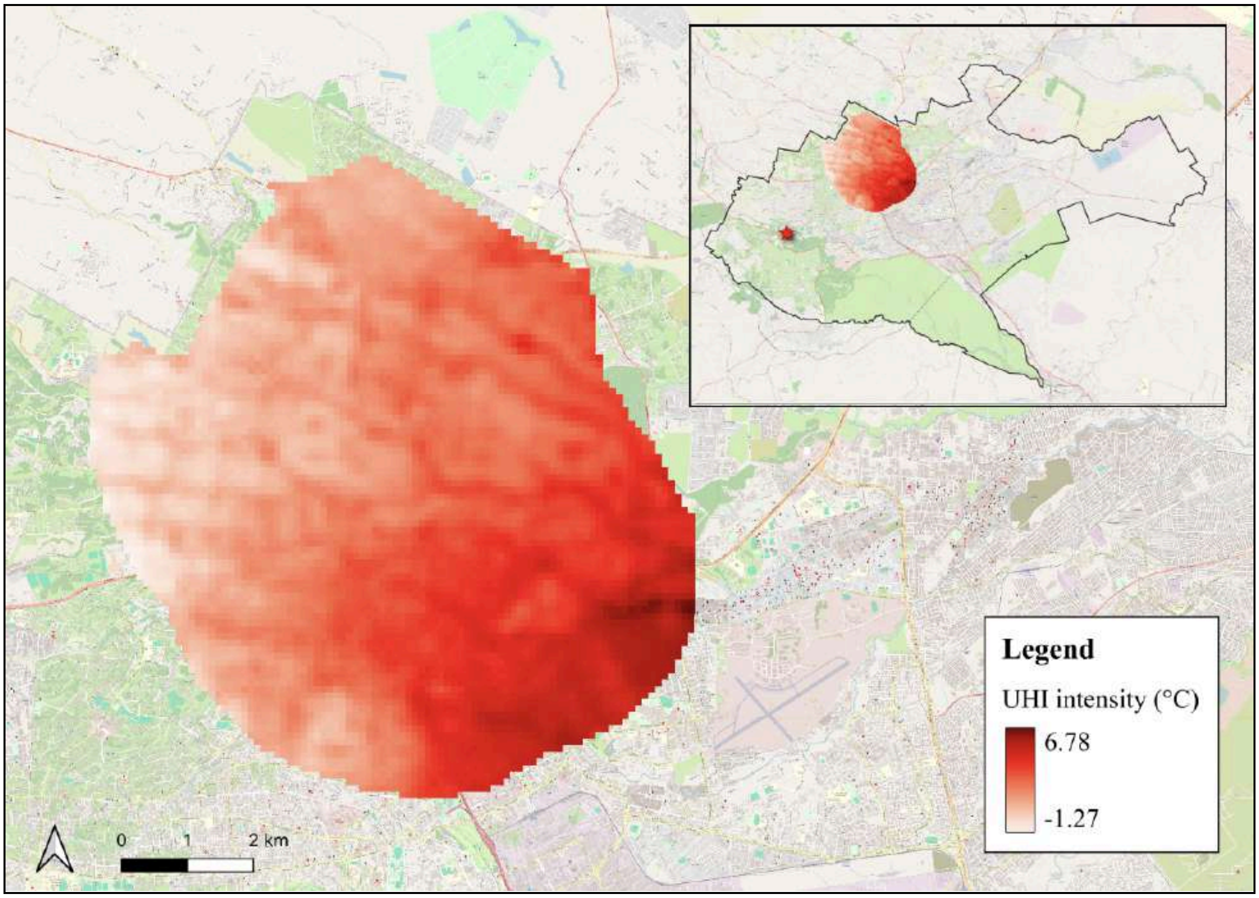

Figure 7: Predicted nocturnal UHI intensity across northern Nairobi based on spatial extrapolation of the September 5th OLS model (elevation and NDVI as predictors). Values represent estimated UHI intensity in °C relative to the Dagoretti rural reference station.

5. Discussion

5.1 UHI Magnitude and Temporal Variation

Mean nocturnal UHI intensity across the four sessions ranged from 1.64°C to 4.81°C, with peak values reaching 6.36°C. These magnitudes are slightly higher than those reported in comparable mobile transect UHI studies in Western cities. Rafiee et al. (2016) recorded UHI intensities ranging from 1 to 3.4°C in Amsterdam during a single summer night. Gunn (2023) found a mean nocturnal UHI of 1.6°C with a maximum of 3.5°C in Inverness across 30 measurement dates. The somewhat higher intensities observed in northern Nairobi suggest that the UHI effect is at least as pronounced as in these Western cities, and possibly more so. This may reflect Nairobi's rapid and largely unplanned urbanisation, previously discussed in the introduction. Variation across sessions likely reflects differences in atmospheric conditions and transect route geometry rather than changes in the underlying urban thermal environment. September 8th, which recorded the lowest mean intensity (1.64°C), had relatively low humidity (51.4%) and moderate wind speed (17.8 km/h). Counterintuitively, September 18th, the windiest and most humid session (24.1 km/h, 83.2%), recorded the highest mean intensity (4.81°C), which might reflect the influence of additional meteorological or environmental factors not captured by the model, or an unusual weather event on that day. This highlights a broader strength of this study: conducting measurements across four sessions, rather than a single night, as done in the previously mentioned studies, allows such anomalous conditions to be identified and accounted for, strengthening confidence in findings that remain consistent across the other sessions.

5.2 The Role of Vegetation

NDVI was the only variable to remain significant across all three accepted sessions in the final SAR models, with SAR coefficients ranging from $\beta = -0.960$ to $\beta = -2.098$ and corresponding OLS coefficients ranging from $\beta = -1.317$ to $\beta = -3.845$. This indicates that vegetation cover was the most stable predictor of nighttime UHI intensity in northern Nairobi. Comparable air-temperature studies using the same methodology remain relatively limited, making direct comparison with regions across the world difficult. However, in Bologna, Italy, Zeynali et al. (2026) identified vegetation proportion (PV), derived from NDVI, as one of the strongest cooling predictors of nocturnal air temperature. Their OLS model produced a PV coefficient of $\beta = -5.99$, while the Spatial Lag Model produced a direct effect of $\beta = -2.56$. Although the Bologna coefficients are generally larger than those observed in Nairobi, both studies consistently demonstrate a substantial cooling influence of vegetation-related variables on nighttime urban temperatures. The stronger Bologna coefficients may partly reflect differences in urban compactness, vegetation distribution, and methodological design, including the use of a vegetation proportion metric rather than NDVI directly. Cristóbal et al. (2008) similarly identified NDVI as one of the strongest predictors of near-surface air temperature across Catalonia, Spain, reinforcing the broader importance of vegetation in regulating urban thermal environments across different climatic and urban contexts.

Given that vegetation cover in northern Nairobi ranges from the dense Karura Forest in the northwest to largely unvegetated commercial areas in the south and east, these results suggest that preserving and expanding green areas is a reliable way to reduce UHI intensity in this part of the city. However, the role of elevation discussed below complicates any simple interpretation based on vegetation alone.

5.3 The Role of Elevation

Elevation was among the strongest predictors of air temperature in the OLS models for September 5th and 8th ($\beta = -0.044$ and -0.018 respectively, both $p < 0.001$), and in the September 5th OLS model it was more significant than NDVI. While elevation is commonly included as a predictor in LST UHI regression studies (de Almeida et al., 2021), there is limited literature for field data, air temperature studies. Nevertheless, previous studies have demonstrated that elevation can represent a significant explanatory factor for urban temperature variability. For example, a study in Zaragoza, Spain, identified elevation as one of the strongest predictors of spatial UHI intensity alongside NDVI, although it used a correlation analysis instead of a regression (Vicente-Serrano et al., 2005).

The September 5th extrapolation raster further illustrates the strong impact of elevation. A point in the south-eastern part of the study area at approximately 1626m elevation showed a predicted UHI of 6.5°C, while a point in the northwest at approximately 1762m showed -0.4°C, a difference of nearly 7°C driven predominantly by a 136m elevation difference. The Karura Forest is also particularly revealing: moving from its eastern to western edge, predicted UHI values shift from approximately 5°C to approximately 1°C across roughly 3km, entirely within a forested area. This means elevation drives substantial thermal variation even in the complete absence of built-up influence. Furthermore, in some parts of the study area, less vegetated locations at higher elevation show lower predicted UHI than densely vegetated areas at lower elevation, further demonstrating that the elevation factor can override the vegetation cooling factor in this context. This finding has important implications for urban planning in high-altitude African cities: interventions focused solely on increasing green space without accounting for the topographical context risks overestimating the cooling achievable through greening in lower-lying areas, while underestimating the natural cooling advantage of higher-elevation locations. Cities like Addis Ababa, situated at around 2400m, likely face similar dynamics worth explicit investigation.

5.4 The Absence of Built-up Variable Significance

Built-up surface, built-up height, and population density did not retain significance in the final SAR models for any session. This contrasts with many LST-based studies in urban contexts. Yin et al. (2018) found building density and impervious surface to be among the strongest predictors of surface UHI in Wuhan using a Spatial Error Model, and Eshetie (2024) found the built-up index to have a strong positive relationship with LST in Addis Ababa ($\beta = 9.332$). Two explanations are most plausible. First, near-surface air temperature integrates heat from a wider surrounding area and is more strongly mediated by atmospheric mixing than LST, which responds directly to the radiative properties of individual surfaces (Sheng et al., 2017; Voogt and Oke, 2003). Second, the transect routes covered predominantly peripheral and upper-CBD areas rather than the densest built-up core of Nairobi. The most intensely built areas such as the lower CBD, industrial zones, and dense informal settlements in the south and east, were not sampled, which likely limited the detectable range of built-up intensity in the dataset.

5.5 Spatial Autocorrelation and SAR Performance

The spatial lag parameters across all three sessions (ρ = 0.712, 0.781, and 0.675) indicate that between 68 and 78% of the spatial variation in UHI intensity at any location is associated with temperatures in the surrounding area. The SAR substantially outperformed OLS in all sessions: pseudo-R^2 values of 0.875, 0.892, and 0.698 for September 5th, 8th, and 18th respectively considerably exceed the corresponding OLS R^2 values of 0.724, 0.690, and 0.522. This mirrors the finding of Zeynali et al. (2026), whose SAR pseudo-R^2 of 0.94 substantially exceeded their OLS R^2 of 0.65 for nocturnal air temperature in Bologna, and is consistent with the general finding of Yin et al. (2018) that spatial regression models outperform OLS when spatial autocorrelation is present in urban temperature data. The similarity of rho values between Nairobi (0.68 to 0.78) and Bologna (0.80) suggests that strong spatial autocorrelation in nocturnal air temperature may be a general feature of urban environments across different climatic and geographic contexts, with important methodological implications: studies modelling urban air temperature using OLS without testing for spatial autocorrelation might risk producing unreliable coefficient estimates and misleading statistical inference.

A related interpretive consideration is that high rho values indicate that the spatial lag captures a substantial share of the spatial variation in UHI intensity, leaving less residual variation for local predictors to explain. When predictors are themselves spatially autocorrelated, as elevation necessarily is because it changes gradually across a landscape, part of their explanatory influence may therefore be absorbed by the spatial lag term rather than appearing directly in the regression coefficients (LeSage & Pace, 2009). This might explain the weakening of the elevation coefficients from OLS to SAR, and should be considered when interpreting SAR coefficients as causal estimates.

To the authors' knowledge, no previous study has applied spatial autoregressive modelling to mobile transect air temperature data in Nairobi. Previous studies in the city relied on correlation analysis, simple OLS, or did not model temperature drivers statistically (Mwangi et al., 2018; Scott et al., 2017). This study therefore represents a methodological step forward for UHI research in East Africa.

5.6 Limitations

Although the study addresses many gaps, it also presents a few limitations. The mobile transect approach is constrained to road-accessible locations and may therefore underrepresent microclimatic conditions in pedestrian-only spaces, interior courtyards, urban parks, or informal settlement environments located away from the road network. The geographic scope is limited to northern Nairobi and excludes the densest built-up areas of the city, which likely contributed to the non-significance of built-up variables. Each session represents a single night, and session-to-session differences in mean UHI intensity should be interpreted cautiously given different route geometries and weather conditions. The one-hour interruption during September 18th may have introduced temporal inconsistency. The high rho values mean the SAR spatial lag may have absorbed real predictor effects alongside spatial clustering, making it difficult to isolate true causal effects from SAR coefficients alone. The uniform 150m neighbourhood radius represents a compromise, ideally, different radii would be optimised separately for each variable type. Finally, at the edges of the study area, focal statistics calculations may have been affected by no-data values beyond the study boundary.

5.7 Future Research

Future studies should extend the transect network to cover a larger portion of Nairobi, including the southern and eastern areas where built-up effects on air temperature are more likely to emerge. Installation of a permanent network of low-cost air temperature sensors distributed across different LCZ types would enable continuous monitoring across seasons. Applying the same spatial regression framework to other high-altitude East African cities would allow comparative analysis of whether the dominance of NDVI and elevation is a consistent regional feature. Combining mobile air temperature measurements with simultaneous LST data, following Zeynali et al. (2026), would enable direct quantification of how surface and atmospheric UHI drivers diverge in Nairobi. Finally, incorporating a larger number of measurement days across different atmospheric conditions and seasons would provide a more comprehensive understanding of the temporal variability and stability of UHI patterns and drivers in the city.

Conclusion

This study investigated the spatial variation and drivers of nighttime UHI intensity across northern Nairobi using mobile air temperature transects and spatial regression modelling. The results demonstrate that nocturnal UHI intensity in northern Nairobi is both substantial and spatially uneven, with mean session intensities ranging from 1.64°C to 4.81°C and peak values exceeding 6°C. Among all tested variables, vegetation cover represented by NDVI emerged as the most consistent and robust predictor of reduced UHI intensity across all accepted sessions, confirming the important cooling role of urban green space within the city. Elevation also exerted a strong influence on local temperature patterns, highlighting the importance of topography in shaping Nairobi's urban temperature.

In contrast, built-up variables such as building height, built surface coverage, and population density did not retain significance in the final SAR models. This suggests that, within the sampled areas of northern Nairobi, vegetation and topography may currently exert a stronger influence on nighttime air temperature than urban density itself. The results further demonstrate the importance of accounting for spatial autocorrelation in UHI studies. SAR models consistently outperformed OLS regression, indicating that urban temperature patterns are strongly spatially structured and that conventional regression approaches may produce unreliable estimates when this structure is ignored.

More broadly, this study contributes to the limited body of UHI research based on air temperature and field data in African cities, and demonstrates the value of combining mobile transects with spatial regression methods in rapidly urbanising contexts. The findings suggest that preserving vegetation cover should represent a central component of urban heat mitigation strategies in Nairobi, while also highlighting the need to consider elevation when interpreting or planning cooling interventions.

Acknowledgments

I would like to thank Kinzah Cordeiro for collecting the field data on which this study is based. I am grateful to Eric Koomen for his guidance and feedback throughout the supervision of this thesis.

Work Cited

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-7799-1

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*. https://doi.org/10.1038/SDATA.2018.214

Cheval, S., Amihăesei, V. A., Chitu, Z., Dumitrescu, A., Falcescu, V., Irașoc, A., Micu, D. M., Mihulet, E., Ontel, I., Paraschiv, M. G., & Tudose, N. C. (2024). A systematic review of urban heat island and heat waves research (1991–2022). *Climate Risk Management*, *44*. https://doi.org/10.1016/j.crm.2024.100603

Cordeiro, K. (2025). *Mobile transect air temperature measurements in northern Nairobi, Kenya* [Unpublished raw data].

Cristóbal, J., Ninyerola, M., & Pons, X. (2008). Modeling air temperature through a combination of remote sensing and GIS data. *Journal of Geophysical Research: Atmospheres*, *113*(D13). https://doi.org/10.1029/2007JD009318

de Almeida, C. R., Teodoro, A. C., & Gonçalves, A. (2021). Study of the Urban Heat Island (UHI) using remote sensing data/techniques: A systematic review. *Environments*, *8*(10). https://doi.org/10.3390/environments8100105

Eshetie, S. M. (2024). Exploring urban land surface temperature using spatial modelling techniques: A case study of Addis Ababa city, Ethiopia. *Scientific Reports*, *14*(1), 6323. https://doi.org/10.1038/s41598-024-55121-6

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gunn, G. F. (2023). Measuring nocturnal near-surface urban heat island intensity in the small, mid-latitude city of Inverness, Scotland. *Scottish Geographical Journal*, *140*(1–2), 192–209. https://doi.org/10.1080/14702541.2023.2242819

Heaviside, C., Macintyre, H., & Vardoulakis, S. (2017). The urban heat island: Implications for health in a changing environment. *Current Environmental Health Reports*, *4*(3), 296–305. https://doi.org/10.1007/s40572-017-0150-3

Howard, L. (1833). *The climate of London, deduced from meteorological observations, made at different places in the neighbourhood of the metropolis*.

Johnson, B. A., & Jozdani, S. E. (2019). Local Climate Zone (LCZ) map accuracy assessments should account for land cover physical characteristics that affect the local thermal environment. *Remote Sensing*, *11*(20), 2420. https://doi.org/10.3390/RS11202420

Kassomenos, P., & Begou, P. (2022). The impact of urban overheating on heat-related morbidity. In *Urban Overheating* (pp. 39–80). https://doi.org/10.1007/978-981-19-4707-0_3

KNBS. (2025). *Kenya National Bureau of Statistics*. https://www.knbs.or.ke/

Kubara, M., & Kopczewska, K. (2023). Akaike information criterion in choosing the optimal k-nearest neighbours of the spatial weight matrix. *Spatial Economic Analysis*, *19*(1), 73–91. https://doi.org/10.1080/17421772.2023.2176539

LeSage, J. P., & Pace, R. K. (2009). Spatial econometric models. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis* (pp. 355–376). Springer. https://doi.org/10.1007/978-3-642-03647-7_18

Li, X., Zhou, Y., Yu, S., Jia, G., Li, H., & Li, W. (2019). Urban heat island impacts on building energy consumption: A review of approaches and findings. *Energy*, *174*, 407–419. https://doi.org/10.1016/j.energy.2019.02.183

Li, L., Zha, Y., & Zhang, J. (2020). Spatially non-stationary effect of underlying driving factors on surface urban heat islands in global major cities. *International Journal of Applied Earth Observation and Geoinformation*, *90*, 102131. https://doi.org/10.1016/J.JAG.2020.102131

Liu, Z., Anderson, B., Yan, K., Dong, W., Liao, H., & Shi, P. (2017). Global and regional changes in exposure to extreme heat and the relative contributions of climate and population change. *Scientific Reports*, *7*(1), 43909. https://doi.org/10.1038/srep43909

Manoli, G., Fatichi, S., Schläpfer, M., Yu, K., Crowther, T. W., Meili, N., Burlando, P., Katul, G. G., & Bou-Zeid, E. (2019). Magnitude of urban heat islands largely explained by climate and population. *Nature*, *573*(7772), 55–60. https://doi.org/10.1038/S41586-019-1512-9

Moede, C. (2021). *Local Climate Zone map of Nairobi* [Dataset]. WUDAPT Community Database.

Mohajerani, A., Bakaric, J., & Jeffrey-Bailey, T. (2017). The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete. *Journal of Environmental Management*, *197*, 522–538. https://doi.org/10.1016/j.jenvman.2017.03.095

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1–2), 17–23. https://doi.org/10.1093/biomet/37.1-2.17

Morris, K. I., Chan, A., Morris, K. J. K., Ooi, M. C. G., Oozeer, M. Y., Abakr, Y. A., Nadzir, M. S. M., Mohammed, I. Y., & Al-Qrimli, H. F. (2017). Impact of urbanization level on the interactions of urban area, the urban climate, and human thermal comfort. *Applied Geography*, *79*, 50–72. https://doi.org/10.1016/j.apgeog.2016.12.007

Muller, C. L., Chapman, L., Grimmond, C. S. B., Young, D. T., & Cai, X. (2013). Sensors and the city: A review of urban meteorological networks. *International Journal of Climatology*, *33*(7), 1585–1600. https://doi.org/10.1002/JOC.3678

Mundia, C. N., & Aniya, M. (2005). Analysis of land use/cover changes and urban expansion of Nairobi city using remote sensing and GIS. *International Journal of Remote Sensing*, *26*(13), 2831–2849. https://doi.org/10.1080/01431160500117865

Mwangi, P. W., Karanja, F. N., & Kamau, P. K. (2018). Analysis of the relationship between land surface temperature and vegetation and built-up indices in Upper-Hill, Nairobi. *Journal of Geoscience and Environment Protection*, *6*(1), 1–16. https://doi.org/10.4236/gep.2018.61001

Nichol, J. E., & Wong, M. S. (2008). Spatial variability of air temperature and appropriate resolution for satellite-derived air temperature estimation. *International Journal of Remote Sensing*, *29*(24), 7213–7223. https://doi.org/10.1080/01431160802192178

Ninyerola, M., Pons, X., & Roure, J. M. (2007). Objective air temperature mapping for the Iberian Peninsula using spatial interpolation and GIS. *International Journal of Climatology*, *27*(9), 1231–1242. https://doi.org/10.1002/JOC.1462

Oke, T. R. (1981). Canyon geometry and the nocturnal urban heat island: Comparison of scale model and field observations. *Journal of Climatology*, *1*(3), 237–254. https://doi.org/10.1002/joc.3370010304

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, *108*(455), 1–24. https://doi.org/10.1002/qj.49710845502

Oke, T. R. (2004). *Initial guidance to obtain representative meteorological observations at urban sites* (IOM Report No. 81). World Meteorological Organization.

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., Carioli, A., Corbane, C., Dijkstra, L., Florio, P., Friedrich, H. K., Gao, J., Leyk, S., Lu, L., Maffenini, L., Mari-Rivero, I., Melchiorri, M., Syrris, V., Van Den Hoek, J., & Kemper, T. (2024). Advances on the Global Human

Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, *17*(1). https://doi.org/10.1080/17538947.2024.2390454

Rafiee, A., Dias, E., & Koomen, E. (2016). Local impact of tree volume on nocturnal urban heat island: A case study in Amsterdam. *Urban Forestry & Urban Greening*, *16*, 50–61. https://doi.org/10.1016/J.UFUG.2016.01.008

Ramírez-Aguilar, E. A., & Lucas Souza, L. C. (2019). Urban form and population density: Influences on Urban Heat Island intensities in Bogotá, Colombia. *Urban Climate*, *29*, 100497. https://doi.org/10.1016/J.UCLIM.2019.100497

Romero Rodríguez, L., Sánchez Ramos, J., Sánchez de la Flor, F. J., & Álvarez Domínguez, S. (2020). Analyzing the urban heat island: Comprehensive methodology for data gathering and optimal design of mobile transects. *Sustainable Cities and Society*, *55*, 102027. https://doi.org/10.1016/J.SCS.2020.102027

Scott, A. A., Misiani, H., Okoth, J., Jordan, A., Gohlke, J., Ouma, G., Arrighi, J., Zaitchik, B. F., Jjemba, E., Verjee, S., & Waugh, D. W. (2017). Temperature and heat in informal settlements in Nairobi. *PLOS ONE*, *12*(11), e0187300. https://doi.org/10.1371/journal.pone.0187300

Sheng, L., Tang, X., You, H., Gu, Q., & Hu, H. (2017). Comparison of the urban heat island intensity quantified by using air temperature and Landsat land surface temperature in Hangzhou, China. *Ecological Indicators*, *72*, 738–746. https://doi.org/10.1016/J.ECOLIND.2016.09.009

Simwanda, M., Ranagalage, M., Estoque, R. C., & Murayama, Y. (2019). Spatial analysis of surface urban heat islands in four rapidly growing African cities. *Remote Sensing*, *11*(14), 1645. https://doi.org/10.3390/RS11141645

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, *93*(12), 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1

Szymanowski, M., & Kryza, M. (2012). Local regression models for spatial interpolation of urban heat island — an example from Wrocław, SW Poland. *Theoretical and Applied Climatology*, *108*(1–2), 53–71. https://doi.org/10.1007/S00704-011-0517-6

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234. https://doi.org/10.2307/143141

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

United Nations. (2019). *World Urbanization Prospects: The 2018 Revision*. United Nations.

University of Cape Town. (2017). *Nairobi climate profile: Full technical version*.

Vicente-Serrano, S. M., Cuadrat-Prats, J. M., & Saz-Sánchez, M. A. (2005). Spatial patterns of the urban heat island in Zaragoza (Spain). *Climate Research*, *30*(1), 61–69. https://doi.org/10.3354/cr030061

Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment*, *86*(3), 370–384. https://doi.org/10.1016/S0034-4257(03)00079-8

Weng, Q., Lu, D., & Schubring, J. (2004). Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment*, *89*(4), 467–483. https://doi.org/10.1016/J.RSE.2003.11.005

Yin, C., Yuan, M., Lu, Y., Huang, Y., & Liu, Y. (2018). Effects of urban form on the urban heat island effect based on spatial regression model. *Science of the Total Environment*, *634*, 696–704. https://doi.org/10.1016/J.SCITOTENV.2018.03.350

Zeynali, R., Mandanici, E., & Bitelli, G. (2026). Urban heat island: Assessing the influence of urban morphology on air and surface temperatures. *Sustainability*, *18*(3), 1695. https://doi.org/10.3390/su18031695

AI Statement:

GenAI was used during the writing of this thesis as a writing and coding assistant. It supported the drafting and editing of text, the development of R code for spatial regression analysis, and the formatting of tables and references. All analytical decisions, interpretations, and conclusions are the author's own.

Annex

Table 2: Summary of Variables, Data Sources, and Resolution Metadata for Urban Heat Island Analysis

Dataset	**Source**	**Year**	**Original Resolution**	**Data Type**
Mobile Temperature Data	PCE Log Tag Analyser (Cordeiro, 2025)	Aug–Oct 2025	–	Point
GPS Coordinates	Strava (Cordeiro, 2025)	Aug–Oct 2025	–	Point
Dagoretti Weather Station	Meteostat (Station ID: 63741)	Aug–Oct 2025	–	Point
Nairobi Boundary	Global Administrative Areas (GADM), version 4.1	2025	–	Vector
NDVI	Sentinel-2 SR Harmonized via Google Earth Engine	Sept 2025	10m	Raster
Elevation	NASA SRTM via Google Earth Engine	2000	30m	Raster
Built-up Surface	GHS Built-up Surface R2023 (Pesaresi et al., 2024)	2025	100m	Raster
Built-up Heights	GHS Built-up Height R2023 (Pesaresi et al., 2024)	2018	100m	Raster
Population Density	GHS Population Grid R2023 (Pesaresi et al., 2024)	2025	100m	Raster
Local Climate Zones (LCZ)	WUDAPT (Moede, 2021)	2021	100m	Raster

Table 7: LCZ Classification (Stweart & Oke, 2012)

Value	Official LCZ Name	Brief Description
1	Compact High-rise	Dense tall buildings (skyscrapers/CBD)
2	Compact Mid-rise	Dense 3–9 story buildings (old city centers)
3	Compact Low-rise	Dense 1–3 story buildings (informal settlements/slums)
4	Open High-rise	Tall buildings with space/trees between them
5	Open Mid-rise	3–9 story buildings with open space (modern apartments)
6	Open Low-rise	1–3 story buildings with space (suburbs/residential)
7	Lightweight Low-rise	Shacks/temporary structures (slums/industrial)
8	Large Low-rise	Large warehouses, factories, malls
9	Sparsely Built	Very few buildings, mostly open land
10	Heavy Industry	Industrial complexes, refineries
11 (A)	Dense Trees	Heavily forested areas
12 (B)	Scattered Trees	Orchards, parklands, savannah
13 (C)	Bush, Scrub	Low woody vegetation, shrubs
14 (D)	Low Plants	Grasslands, agricultural fields
15 (E)	Bare Rock/Paved	Parking lots, rocky outcrops
16 (F)	Bare Soil/Sand	Fallow fields, construction sites, deserts
17 (G)	Water	Lakes, rivers, large ponds

Source: Stewart and Oke (2012).